PROGRAMMED METHYLATION AND EDITING IN FUNGAL POLYKETIDE BIOSYNTHESIS

by
Philip A. Storm

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy in Chemical Biology

Baltimore, Maryland
October 2017

Abstract

Fungal polyketides are natural products assembled by polyketide synthases (PKSs), and among this rich class of secondary metabolites are potent toxins, pigments, and best-selling pharmaceuticals. Such diverse bioactivities all originate from the same core biosynthetic reactions. PKSs catalyze the condensation and modification of small acyl substrates to generate complex scaffolds in an iterative and programmed fashion. Substrates and intermediates are tethered to the PKS and delivered to client domains that act repeatedly during the synthesis of a single product molecule. Non-reducing PKSs (NR-PKSs) generate linear poly-β-ketone intermediates of a specific chain length that are subsequently cyclized in a regiospecific fashion and released to give aromatic mono- and polycyclic products, many of which are further modified by downstream enzymes.

In the last decade, significant effort in the Townsend lab has been devoted to understanding NR-PKS programming, a challenging task given the size of these megasynthases and the lack of isolable intermediates in the native producers. Previous work pioneered a domain deconstruction approach, whereby NR-PKSs are dissected at the genetic level to generate mono- and multidomain fragments through heterologous expression. Reconstituting these fragments *in vitro* enables examination of specific domain combinations, shedding light on the mechanisms and programming of individual catalytic domains. Domain deconstruction has also provided opportunities to obtain atomic-resolution structures of NR-PKS domains.

Some NR-PKSs contain an embedded *C*-methyltransferase (CMeT) that installs one or more methyl groups onto the polyketide scaffold, but little was known about the mechanism or programming of this enzymatic alkylation. In this work, we extend the

domain deconstruction approach to explore NR-PKS *C*-methylation, particularly in the biosynthesis of citrinin, a fungal toxin. In addition to biochemical experiments designed to tease apart how different CMeTs install different methylation patterns, we solved the first crystal structure of a NR-PKS CMeT and found characteristic features that are consistent across all known PKS CMeTs, from marine bacteria to pathogenic fungi. Finally, we identified an editing hydrolase in gene clusters with CMeT-containing NR-PKSs and propose that it serves to maintain efficient biosynthesis by removing off-path intermediates from the PKS *in trans*. The work in this dissertation will serve as a foundation to rationally engineer methylation and editing of polyketide biosynthesis.

Thesis Advisor:

Dr. Craig A. Townsend, Department of Chemistry, Johns Hopkins University

Additional Readers:

Dr. Dominique Frueh, Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine

Dr. Marc Ostermeier, Department of Chemical and Biomolecular Engineering, Johns Hopkins University

Dedicated to my parents

Dean and Dorothy

For their constant love and support

And though the course may change sometimes

Rivers always reach the sea

- Led Zeppelin, “Ten Years Gone” -

Acknowledgements

I have had the great privilege of befriending and working with talented scientists and wonderful people during my time at Johns Hopkins University. And I have relied upon family and friends during moments of despair and jubilation.

Prof. Craig Townsend recruited me to Hopkins, and by bringing me into his lab, has enabled me to grow and develop as a scientist. His steady hand and earnest guidance have led me to develop skills that I thought were well outside my aptitude, and I am deeply grateful for his mentorship and tireless efforts captaining the lab. I have been fortunate to be trained by brilliant senior students, especially Dr. Anna Vagstad and Dr. Adam Newman, both of whom contributed significantly to my understanding of biology and biochemistry. They set a standard of excellence that has served as a guiding star. Fellow members of Team PKS, including Dr. Callie Huitt-Roehl, Doug Cohen, Jacob Kravetz, Jamie Alley, and Paramita Pal have been incomparable colleagues, and I will remember our collaborative moments fondly – we have and will continue to advance PKS knowledge considerably. Other members of the lab have been dear friends and always quick to offer advice, or a cold beer when necessary, especially Dr. Rongfeng Li, Dr. Jesse Li, Dr. Courtney Hastings, Dr. Victor Outlaw, Dr. Dan Marous, Dr. Evan Lloyd, Ryan Oliver, and Dr. Darcie Long. I know I made the right decision to come to Baltimore because I've shared my graduate school experience with all of you.

I'd also like to extend my thanks to our collaborators. Prof. Timm Maier and Dr. Dominik Herbst began working with the Townsend Lab at a pivotal moment in my time here, and their efforts and insights have greatly augmented our own. I eagerly await the future fruits of this collaboration and expect many accolades for both of them. I'd also

like to thank Jurgen Bosch and Lauren Boucher for their crystallography expertise at a time when I lacked any of my own.

Through Hopkins, I've been advised and assisted by fantastic faculty and support staff. Thanks to Prof. Steve Rokita and Prof. Marc Greenberg for their leadership of the CBI program, and again to Prof. Rokita and Prof. Caren Meyers for their honest appraisals and thoughtful suggestions as members of my thesis committee. I also want to thank Prof. Dominique Frueh and Prof. Marc Ostermeier for serving as readers for my thesis on short notice. Dr. Phil Mortimer and Dr. Michael McCaffery have been invaluable resources for mass spectrometry and imaging tools, respectively. Without Lauren McGhee, the CBI program would have suffered a fledgling death, and her ability to navigate the graduate school process for so many students is a superpower. Additionally, Boris Steinberg has kept the roof above our heads, ensured the liquid nitrogen was flowing, and always brought expert craftsmanship to bear. I owe all of you for maximizing my experience at Hopkins.

I also owe a great debt to my loved ones outside of science, who have always bolstered my spirits and allowed me to remember why such hard work is worth the time and effort. Patrick Owlett and Jake Christopher have been dear friends and their arrival in Baltimore has extended the fun and frivolity of our college days. I'm very excited to witness his marriage to Annie next spring – he's chosen wisely. Of course, my parents, Dean and Dorothy, have made my path in life possible. Their sacrifices and unwavering support of my education and research can never be repaid, but I have been incredibly fortunate to share my accomplishments with them. I have been fortunate to be an only child as "my cup runneth over."

Finally, I want to praise the love and support of Callie Huitt-Roehl as my partner in both science and life. Without her compassion, intellect, and culinary prowess, I doubt I could have come this far. I've never known such a compatible companion, and I can't wait for our next challenges together.

Publications

Storm, P.A. and Townsend, C.A. Exploring fungal polyketide C-methylation through combinatorial domain swaps. 2017, *In preparation*.

Storm, P.A. and Townsend, C.A. *In trans* hydrolysis of carrier protein-bound acyl intermediates by CitA during citrinin biosynthesis. 2017, *under review, Chem. Comm*.

Storm, P.A., Herbst, D.A., Maier, T., and Townsend, C.A. Functional and Structural Analysis of Programmed C-Methylation in the Biosynthesis of the Fungal Polyketide Citrinin. *Cell Chem. Biol*. 2017, **24**, 316-325.

Newman, A.G., Vagstad, A.L., Storm, P.A., and Townsend, C.A. Systematic Domain Swaps of Iterative, Nonreducing Polyketide Synthases Provide a Mechanistic Understanding and Rationale For Catalytic Reprogramming. *J. Am. Chem. Soc*. 2014. **136**, 7348-7362.

Vagstad, A.L., Newman, A.G., Storm, P.A., Belecki, K., Crawford, J.M., and Townsend, C.A. Combinatorial domain swaps provide insights into the rules of fungal polyketide synthase programming and the rational synthesis of non-native aromatic products. *Angew. Chem. Int. Ed. Engl*. 2013, **52**, 1718-1721.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to polyketides and programmed biosynthesis

1.1. Introduction to natural product biosynthesis

Almost every crevice on the face of the planet has been found to be home to microbial life.[1] It is a testament to the power of evolution by natural selection that organisms can occupy such diverse realms only to find stiff competition for survival. Like all life, microbes take in nutrients and produce molecules necessary for survival through metabolic pathways – interconnected collections of genes, proteins, and chemical reactions. Primary metabolism includes the pathways necessary to produce essential metabolites; amino acids for protein biosynthesis, lipids for membrane formation and energy storage, and nucleic acids for replicating genetic information are all examples of primary metabolites.[2] A deficiency in one of these pathways is frequently detrimental, and often lethal, to the organism.

Microbes are also an astonishingly rich source of secondary metabolites (Figure 1.1).[3] Broadly, these are molecules not essential for basic cellular functions but which provide a survival advantage. Secondary metabolites serve their producers by inhibiting growth of competitors, scavenging scarce nutrients, facilitating infection of a host, or transmitting information to neighboring cells.[4] Mankind has been using naturally-derived substances rich in secondary metabolites for millennia, generally for therapeutic effects like the pain relief associated with chewing a particular tree's bark. Only for the last century or so, have microbes been exploited for their wealth of secondary metabolites, initially through the fermentation and extraction of isolated strains of bacteria or fungi.

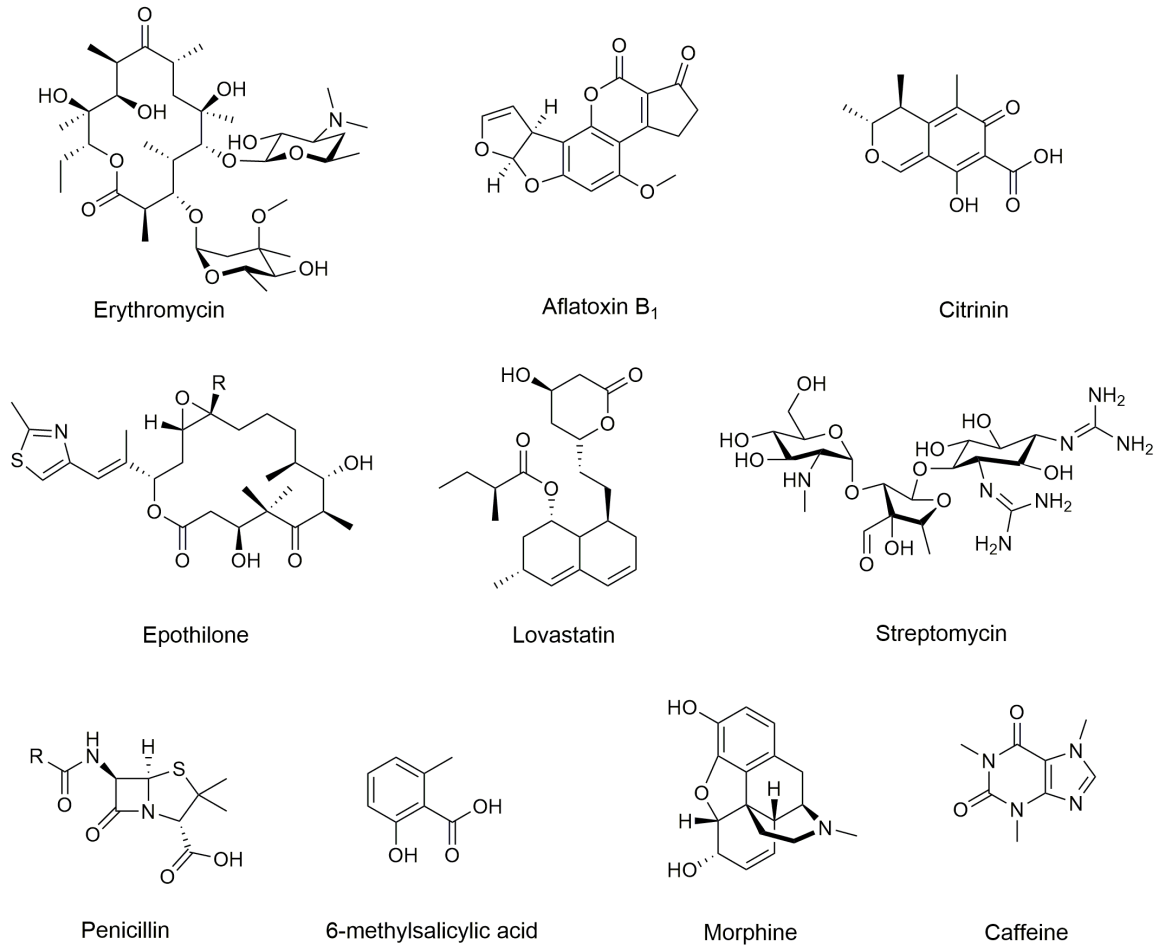


Figure 1.1: Examples of natural products from several classes. Microbial natural products are as diverse in their bioactivities as in their structures. Erythromycin, streptomycin, and penicillin are antibiotics, whereas aflatoxin and citrinin are potent toxins for humans and livestock. Lovastatin led to the development of anti-hypercholesteremia drugs and morphine and its derivatives are widely (and perhaps overly) used for pain management.

Early efforts led to the development of several classes of antibiotics, cholesterol-lowering statins, immunosuppressants, and chemotherapeutics.[5] As technology has advanced, the scope of microbial natural products has widened. Genomic efforts have not only identified the genes and proteins responsible for the production of known molecules, but also have illuminated a greater capacity for biosynthesis than had been appreciated.[6, 7] Fortunately, the genes of a given biosynthetic pathway in bacteria and most fungi are almost always physically clustered on the chromosome, enabling bioinformatic

applications like antiSMASH to rapidly identify potential biosynthetic pathways and predict their products.[8] The advantages of clustering include common regulation and expression of all genes in the cluster and ease of horizontal transfer.[9] The latter is thought to be partially responsible for the diversity of natural products, as clusters can be rapidly duplicated, spread, and diversified across many organisms, with each retaining the variants most suitable for its own ecological niche.[10] Furthermore, metagenomic analyses have cataloged the biosynthetic pathways found in specific environments without the need for axenic cultures or even growth under laboratory conditions, a bottleneck that otherwise restricts ~99% of microbes from study.[11, 12]

The overwhelming amount of data, and the number of unexplored biosynthetic pathways, has renewed interest in exploring natural products as a source of therapies.[13] As a group, these molecules tend to have complex scaffolds often difficult to replicate synthetically and occupy regions of chemical space likely enriched for biological activity. In fact, 0.3% of known polyketide scaffolds had resulted in a drug as of 2005, a 300-fold enrichment over traditional pharmaceutical screens.[14] Their biosynthetic pathways also have a number of features that make them attractive for engineering efforts, inspiring thoughts of "plug-and-play" cassettes that can be used to produce novel products cleanly and efficiently.[15] How researchers can best co-opt these biosynthetic pathways remains an open question, however newer tools and technologies are poised to deepen our understanding of the "programs" that biosynthetic enzymes follow to make a given product. The work presented in the following chapters is a contribution to untangling the programmed biosynthesis of fungal polyketides, particularly those that are *C*-methylated.

1.2. Polyketides and iterative catalysis

Three major classes of natural products share the same core biosynthetic logic: polyketides, fatty acids, and non-ribosomal peptides.[16, 17] In each case, the core scaffold of the natural product is built through sequential rounds of condensation and modification of simple substrates, and reactions are carried out on intermediates covalently bound to a carrier protein (Figure 1.2). Polyketide synthases (PKSs) and fatty acid synthases (FASs) include acyl carrier proteins (ACPs) while non-ribosomal peptide synthetases (NRPSs) use a peptidyl carrier protein. A post-translational modification of a conserved serine residue to form a phosphodiester with the phosphopantetheine moiety of Coenzyme A (CoA) converts *apo*-ACP to *holo*-ACP, the active form of the carrier protein.[18] This modification occurs at one end of a small four-helix bundle, giving the carrier protein a flexible arm, approximately 20 Å long. The phosphopantetheinyltransferases (PPTases) that catalyze the modification are found as both discrete enzymes or as embedded domains in larger proteins, often the same protein that contains the ACP. Substrates and intermediates are covalently bound through a thioester bond with the free thiol on the arm and can be delivered into the active sites of other enzymes or domains. Tethering these species to the carrier protein also greatly increases their effective concentration, preserving the metabolic investment used to synthesize complex intermediates.

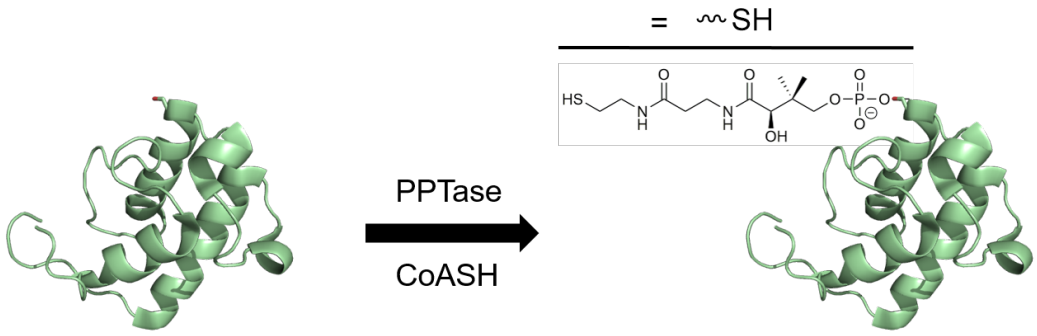


Figure 1.2: Activation of an ACP with phosphopantetheine. The solution structure of the excised ACP monodomain from PksA is shown being activated by a non-specific PPTase to install the phosphopantethein arm.[19] In schematic representations, the arm is often signified with a short, wavy line ending with a thiol or acyl thioester. PPTases are Mg^{2+}-dependent, and several are promiscuous allowing for the loading of synthetic analogs, a valuable tool for probing natural product biosynthesis.

FASs and PKSs share the same set of reactions for condensation and modification[20], in addition to carrier protein-tethered intermediates (thorough discussion of NRPSs is available elsewhere[21]). The sequence of reactions can be carried out by an enzyme with multiple catalytic domains encoded in its primary sequence (Type I) or multiple discrete enzymes that form a complex in solution (Type II).[22, 23] Type II FASs and PKSs are found in bacteria, while Type I multienzymes are present in bacteria, fungi, and animals and include the PKSs used in this work.[24] Small acyl intermediates are loaded onto specific acyltransferase (AT) domains and then loaded onto the ACP for transfer to a conserved cysteine in the ketosynthase (KS), the workhorse domain of PKSs and FASs. The ACP is then loaded with a malonyl extender unit, or derivative, which it delivers to the KS where decarboxylative Claisen condensation occurs. The ACP-bound nucleophile attacks the KS-bound acyl intermediate, resulting in a head-to-tail assembly of two-carbon units per round of extension.

The fate of the newly extended acyl intermediate depends on the presence of tailoring domains in the PKS/FAS and the programmed pathway (Figure 1.3). The most

common modification or tailoring step is reduction of the β-ketone, either to the corresponding alcohol by a ketoreductase (KR), the α,β-unsaturated thioester by a dehydratase (DH), or the completely reduced methylene by an enoyl reductase (ER). FASs rest at one end of a spectrum of reduction; they generally reduce the β-ketone following each round of extension to generate a completely saturated fatty acid, with fungal and bacterial PKSs ranging from completely non-reducing to partially and highly-reducing.[25]

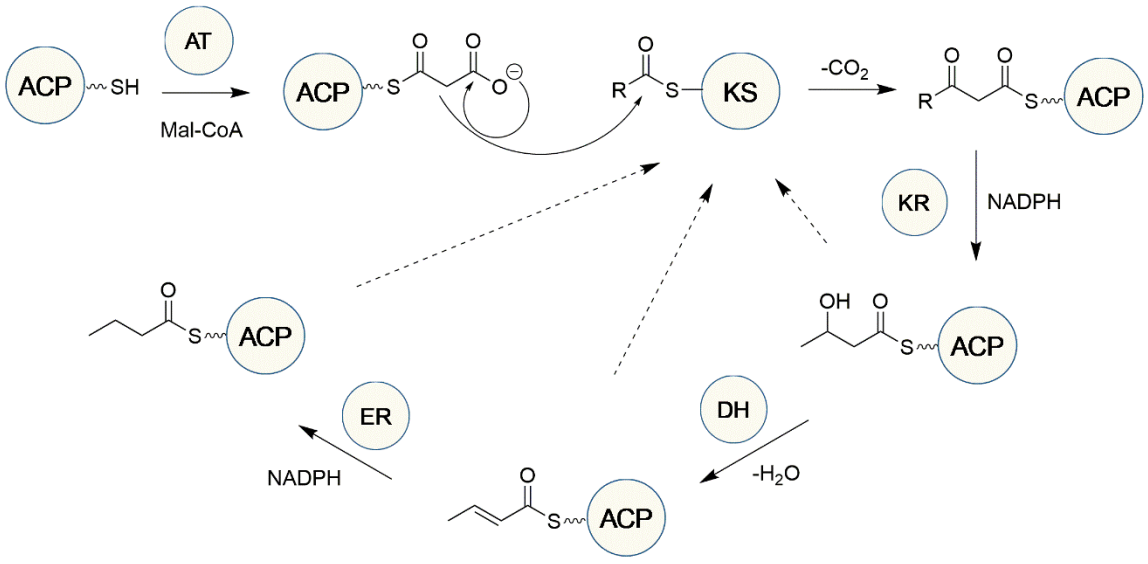


Figure 1.3: Schematic of FAS and PKS condensation and tailoring. The AT domain loads extender substrates onto the *holo*-ACP, which when delivered to the KS allows for decarboxylative Claisen condensation. The resulting substrate, now two carbons longer, can go through all, some, or none of the β-reductions, depending on available domains. In FASs, the entire suite of reductive domains is used, in HR-PKSs, only programmed domains are used, and in NR-PKSs, none are used as none are present, and the extended substrate is loaded back onto the KS for another round of condensation.

The architecture and function of the condensing and modifying domains can be further categorized as modular or iterative. Bacterial modular PKSs (modPKSs) are comprised of multiple "modules" or sets of catalytic domains that complete a single round of condensation and modification before handing off the intermediate to a downstream module, like an assembly line.[14] The sequence of the first modular PKS,

responsible for erythromycin biosynthesis, serves as a canonical example.[26] A single gene can encode one or more modules, and ordering the modules from *N*- to *C*-termini reveals a relationship between domain architecture and product termed co-linearity. Because synthesis also proceeds from *N*- to *C*-termini, one can predict the PKS product by inspecting the modifying domains present in each module (Figure 1.4), however there are known deviations from co-linearity.

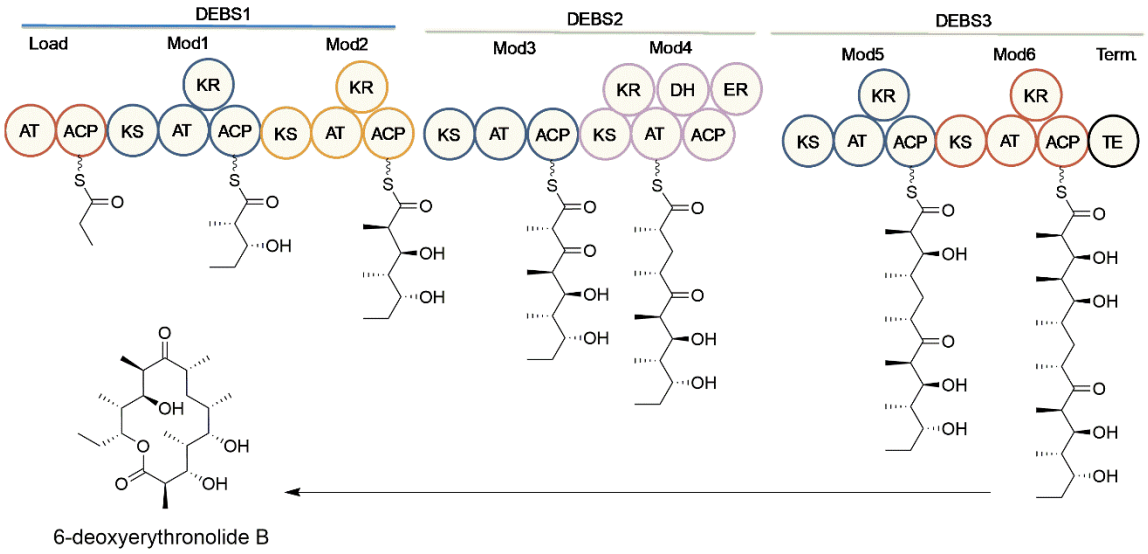


Figure 1.4: Biosynthesis of the erythromycin aglycone is the canonical example of co-linearity in modPKSs. Shown as a "spaghetti" diagram, each module's ACP has the intermediate present upon completion of that module's cycle. Correlating reductive domains with the functionality at the β-carbon is textbook co-linearity. The final linear intermediate is macrolactonized by the TE to release the aglycone for downstream oxidation and glycosylation, ultimately yielding erythromycin.

Upon elucidation of their domain architecture, modPKSs were obvious targets for engineering. Initial efforts perturbed biosynthesis by inactivating a single tailoring domain of a module, for example a KR, and looking for retention of the previously reduced ketone.[27] Swapping of entire modules has also been accomplished[28], inspiring the term "combinatorial biosynthesis," but these efforts have not been as facile as first hoped.[29, 30] Modules connect through *N*- and *C*-terminal helical bundles, and while

careful selection of these can improve hybrid modPKS production, structural data at the multi-module level will be necessary to achieve productive and general combinatorial regimes.[31]

A notable subclass in polyketide biosynthesis is found in the more recently discovered *trans*-AT PKSs.[32] These systems are uncommon in the best studied bacterial polyketide producers, *Actinomycetes*, but may represent over a third of all modPKSs; they are most commonly found in unusual or uncultured branches of the bacterial family tree, particularly symbionts, pathogens, and marine-dwelling bacteria.[33] Even before *trans*-AT polyketides were understood to be a separate class, products like mupirocin were successfully commercialized as antibiotics used to treat drug-resistant infections.[34] As the name implies, these modPKSs lack AT domains within the modules and instead use one or a few discrete AT domains *in trans* to supply each module with the required extender substrates. To add to their peculiarity, the types of *trans*-AT modules (several dozen and counting) greatly outnumber those found in canonical *cis*-AT modPKSs (~8) and are so rich in uncharacterized domains lacking any sequence homology that many modules cannot be assigned a function. Among those that have been characterized, activities include O-methylation, β-γ dehydration, oxygen insertion, and β-branching.[32, 35, 36] *Trans*-AT PKSs most certainly contain yet more undiscovered biosynthetic novelties and deserve significant attention moving forward.

Iterative PKSs (iPKSs) can be found in bacteria, fungi, and even humans (accounting for mammalian FAS). They differ from their modular cousins by reusing the same set of catalytic domains for each round of extension and modification. Unfortunately, this behavior makes predicting the PKS product very difficult as the

number of condensations and the subsequent degree of modification is obscured by their iterative function, as not all domains may be used every cycle. Deciphering the mechanisms by which programmed biosynthesis is enforced remains a primary goal of the field, even though several iPKSs have been extensively studied.[37-40]

1.3. Fungal non-reducing polyketide synthases

1.3.1 Overview

Non-reducing PKS (NR-PKS) products were some of the earliest isolated polyketides, but the basis for their biosynthesis was unknown until relatively recently, even though non-reduced polyketides are of considerable concern as human or agricultural toxins.[20, 41] Aflatoxins, a collection of products whose biosynthesis is dependent on the NR-PKS PksA, are hepatotoxic and carcinogenic compounds produced by a number of filamentous fungi, notably *Aspergillus* species.[42] Infection of crops by aflatoxin-producing species results in accumulation of these mycotoxins that can induce acute or chronic harm to livestock and humans that consume these foods.[43] Rigorous monitoring can reduce the risk of aflatoxin contaminated products reaching the food supply, but many regions of the world lack the resources or infrastructure to carry out such monitoring. Citrinin, one of the oldest known polyketides, also poses a human health risk. It is a nephrotoxin commonly found in fermented rice products, as the *Monascu*s and *Aspergillus* species that are used to produce red pigments during fermentation often co-produce citrinin.[44] The NR-PKS-derived phytotoxin cercosporin is also of particular relevance to agriculture. *Cercospora* species infect dozens of high-value

crops, including beats, soybeans, coffee, and tobacco, causing a characteristic "leaf spot."[45] Cercosporin is exported from the fungus and serves as a photoactivator that converts oxygen to the singlet state, which oxidizes plant membranes to cause nutrient leakage and enables further fungal pathogenicity.[46]

Applying genetic, biochemical, and structural techniques to probe iPKS programming has been a longstanding effort in the Townsend lab, particularly in the context of fungal NR-PKSs. Identifying and characterizing the NR-PKSs responsible for toxin production can serve efforts to limit contamination and crop loss, and the dozens of reactions catalyzed during NR-PKS biosynthesis offer a number of targets for engineering biosynthesis.

1.3.2. Unveiling NR-PKS programming through domain dissection

The first breakthrough in efforts to explore NR-PKS biosynthetic programming came with the discovery that domain boundaries could be identified through bioinformatics. The Udwary-Merski algorithm combines multiple sequence analysis with predicted secondary structure and local hydrophobicity to score each residue in the megasynthase (Figure 1.5).[47] High scores indicate that a position is conserved, in a α-helix or β-sheet, or likely buried in the core of a domain, but low scores reflect interdomain linker regions that are likely to be flexible. With the domain boundaries identified, sequences of mono- or multidomain fragments can be inserted into heterologous expression vectors. The method was first applied to PksA, the aflatoxin NR-PKS from *A. parasiticus*, and led to the first *in vitro* reconstitution reactions of a NR-PKS.[37]

Domain dissection was then applied to several additional NR-PKSs, including CTB1 (cercosporin, *C. nicotianae*[48]), Pks1 (tetrahydroxynaphthalene, *Colletotrichum lagenarium*[49]), Pks4 (bikaverin, *Gibberella fujikuroi*[50]), wA (melanin, *A. nidulans*[51]), and ACAS (atrochrysone carboxylic acid, *A. terreus*[52]). Though not every domain fragment could be obtained for each NR-PKS, a sufficiently large collection was assembled to conduct combinatorial domain swaps.[53, 54] The following sections describe the lessons learned from these experiments and others, and serve as the foundation for the work described in subsequent chapters.

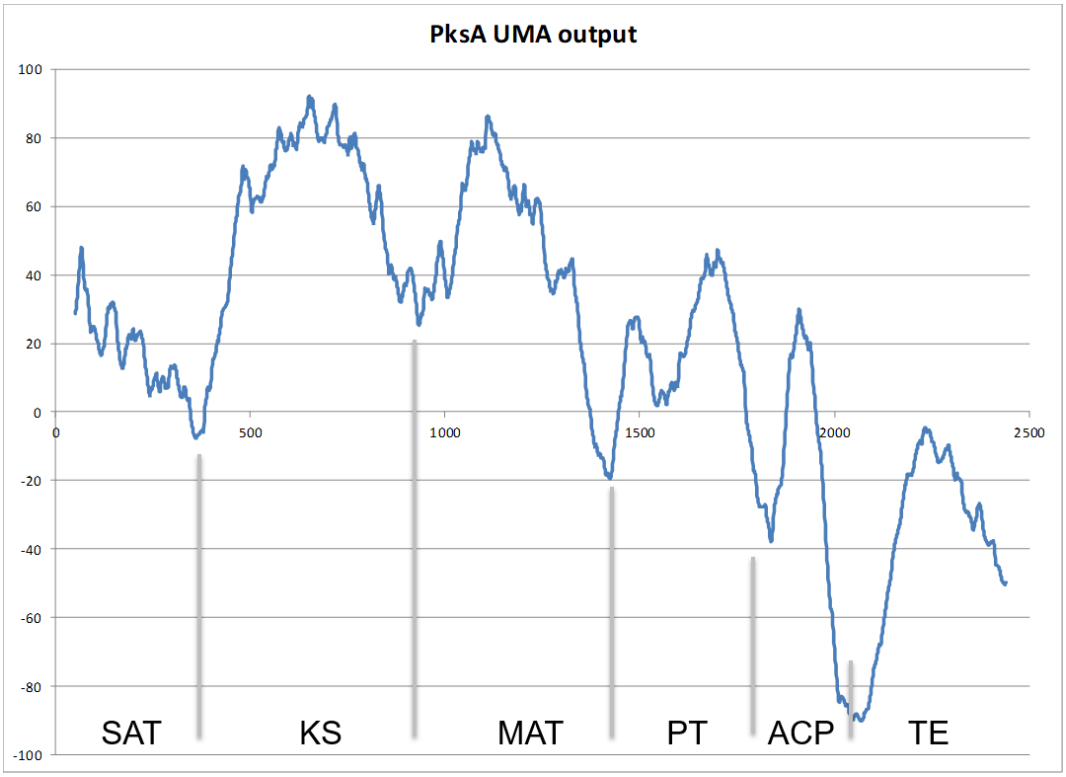


Figure 1.5: Example output from the Udwary-Merski algorithm. Each residue in PksA is represented along the x-axis, with the residue's score on the y-axis. The domain boundaries are marked by the gray lines. The absolute value of the score is less predictive than the local maxima and minima due to the varying levels of conservation from domain to domain. For example, KS domains are highly conserved across all PKSs, whereas lower conservation of SAT domains in NR-PKSs can be attributed to the different starter units selected (or presence of an upstream starter unit supplier). Using the local minima to guide the selection of domain boundaries allows for heterologous expression of mono- and multidomain fragments.

1.3.3. Initiation and Extension

NR-PKSs share the essential domains for decarboxylative Claisen condensation – KS, ACP, and malonyl-CoA:ACP acyltransferase (MAT) – common to all PKSs, but also contain unique domains. In modPKSs, the acyl species that serves as the electrophile in the first condensation reaction, termed the starter unit, is generally loaded onto the enzyme by an intiation module.[31] This small module consists of either a loading AT or KS^Q domain (a mutated KS capable of decarboxylation but not condensation) paired with its own ACP that transfers the starter unit to the downstream module's KS for condensation.[55] Highly-reducing (HR-) and partially-reducing PKSs (PR-PKSs) have no *N*-terminal loading domain, but instead the starter unit is directly loaded onto the KS from the corresponding acyl-CoA.[56, 57]

NR-PKSs have a dedicated starter unit:ACP transacylase (SAT) that selects and loads the starter unit onto the ACP (Figure 1.6).[58] Such a specialized domain had been hypothesized on the basis of radiolabel studies showing that acetyl-CoA was used directly rather than through decarboxylation of malonyl-CoA.[59] Previous work in the Townsend lab identified an *N*-terminal domain of unknown function with sequence similarity to other ATs that selectively transferred starter units to the ACP.[60] Many NR-PKSs use more elaborate starter units that are provided by a dedicated upstream FAS or HR-PKS, and more recent work has suggested that the SAT domain also serves as a protein adapter to facilitate transfer from the upstream ACP to the NR-PKS ACP.[61]

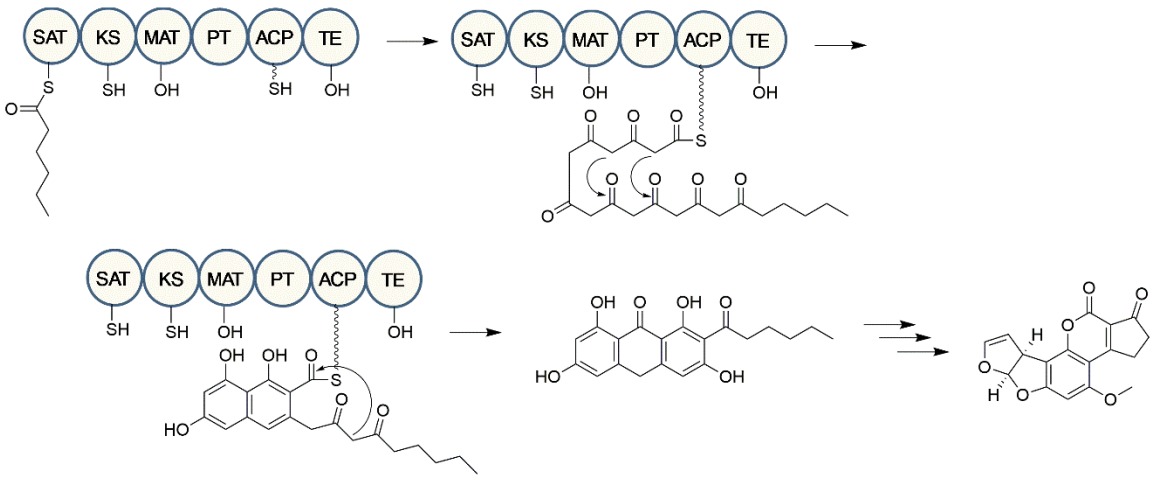


Figure 1.6: Representative example of NR-PKS biosynthesis. PksA accepts a hexanoyl starter unit from a dedicated pair of yeast-like FASs that form a complex with PksA.[62] The starter unit is transferred to the KS, and seven rounds of iterative condensation produce a C_{20} poly-β-ketone intermediate that is regiospecifically aldol cyclized from C4-C9 and C2-C11 by the PT domain. The bicyclic intermediate is released by TE-catalyzed Claisen cyclization to the noranthrone. A series of complex oxidative rearrangements results in aflatoxin B_1.

Once the starter unit is transferred to the KS, condensation to the full chain length is rapid and processive, with no detectable intermediate chain lengths.[63] The growing intermediate is likely retained in the active site throughout by rapid transfer from the ACP back to the KS following condensation. Combinatorial swaps indicate that chain lengths are controlled predominantly by the KS until the intermediate approaches the programmed length, at which point downstream domains, particularly the product template (PT), appear to participate as well.[54] In the absence of any downstream domains a variety of derailment products are observed. These species tend to be full-length intermediates that spontaneously cyclize to a number of products also seen in other examples of non-reduced polyketide biosynthesis.[54, 64] After the triketide stage, intermediates can also spontaneously release themselves from the ACP thioester by C-O bond formation yielding a pyrone.

1.3.4. Regiospecific cyclization and release

The second unique feature of NR-PKSs is the PT, which catalyzes regiospecific aldol cyclization of the full-length ACP-bound poly-β-ketone intermediate to give a mono- or bicyclic intermediate, with loss of one or two water molecules.[37] The PT domain, while not similar by sequence, adopts a double hot dog fold resembling the fold found in DH domains in other iPKSs, consistent with the formal dehydration carried out during PT-mediated cyclization.[65] Three common cyclization patterns have been observed giving rise to monocyclic C2-C7, and bicyclic C4-C9/C2-C11 and C6-C11/C4-C13 intermediates. An additional non-natural C8-C3 pattern has been induced through mutation of a PT domain.[66]

Once cyclized, the acyl-ACP intermediate is a substrate for thioesterase (TE)-catalyzed release. TEs operating in PKS and NRPS biosynthesis have a number of functions beyond canonical hydrolysis. NR-PKS TE domains have been shown to release products through Claisen cyclization, lactonization, and hydrolysis, and Pks1 TE also has deacetylase activity.[67-69] Additionally, TEs often serve a role in editing non-productive or off-path intermediates. For example, PksA TE hydrolyzes hexanoyl intermediates from the ACP but not malonyl extender units, presumably to prevent stalling of the PKS in the event of malonyl decarboxylation without extension or improper loading of hexanoyl to the ACP instead of malonyl or the growing KS-bound intermediate.[63]

1.3.5. Unexplored NR-PKS domain functions

The domain deconstruction and *in vitro* reconstitution reactions discussed above provided a set of "rules" for how starter unit selection, extension, cyclization, and release were carried out. Other aspects of NR-PKS biosynthesis were yet to be examined at the

outset of the work presented herein. The most common NR-PKS domain not included in the initial combinatorial swaps was a putative *C*-methyltransferase (CMeT), which can be found in up to a third or more of NR-PKSs, depending on the organism.[70] These domains install one or more methyl groups during chain elongation, but the mechanism and programming of methyltransfer were not clear at the outset of this work. That NR-PKS *C*-methylation remained unexplored despite representing the simplest example of programming emphasized the difficulty of characterizing individual intermediates in iPKSs. Expanding domain deconstruction and combinatorial domain swaps to explore how *C*-methylation is programmed was a logical extension of previous work.

Some NR-PKSs also contain a *C*-terminal reductase (R) domain in place of a TE, thought to catalyze NAD(P)H-dependent hydride transfer that reduces the acyl-ACP thioester to an aldehyde product and free *holo*-ACP.[70] The first R domains were identified in yeast lysine biosynthesis and later in the NRPS-based biosynthetic pathway for myxochelins.[71-73] In those cases, both two and four electron reductions are catalyzed, though such secondary reductions in polyketide biosynthesis are not known at this time. Though relatively few examples of NR-PKS R domains are known, their aldehyde products often serve as handles for subsequent cyclization or transamination.[74-76] No known examples of R-catalyzed editing are known, however. The fundamental question of how R-containing NR-PKSs remove off-path intermediates was also of particular interest as the project developed.

1.4. Structural lessons and future directions in natural product biosynthesis and engineering

Despite the phenomenal expansion in knowledge about polyketide biosynthesis during the last two decades, many questions remain, particularly surrounding the mechanisms of programming and their structural basis. Crystal structures of individual domains have clarified their folds and active site organizations, and some structural information – either direct of through homology – is available for almost all NR- and HR-PKS domains. Early achievements for iPKS structural fragments included a DEBS KS-AT didomain, the PksA PT and TE, and several β-tailoring domains.[65, 67, 77-80] Many of these structures include bound substrates or substrate mimics, but not all substrates are amenable to crystallographic studies, especially the highly-reactive poly-β-ketone intermediates found in NR-PKSs though some pseudo-substrates have been visualized.[81]

Monodomain structures have been incredibly informative when addressing catalytic mechanisms of individual domains, but they do not capture the larger scale mutlidomain organization or conformation changes in iPKSs. Crystallography has been used to investigate cross-linked structures where an ACP is covalently linked to a client domain in both FAS and PKS examples, however, demonstrating consistent themes in ACP:client domain interfaces.[82, 83] The landmark crystal structure of mammalian FAS by Maier, Leibundgut, and Ban provided the first thrilling glimpse into the three-dimensional organization of PKSs.[84] The megasynthase was revealed to be a head-to-head dimer with primary interfaces between the KS and DH domains, affectionately referred to as a "gingerbread man." The ACP was not visible, however, as a result of its inherent mobility and the lack of any tools to force docking at a single client domain.

Maier and coworkers recently published the composite structure of a HR-PKS involved in mycocerosic acid biosynthesis in *Mycobateria* that revealed conformational coupling in the β-tailoring domains, leading to a proposal for how programmed processing occurs.[85]

The most relevant development in structural biology for addressing complete PKSs has been the increasing accessibility and resolution of cryo-electron microscopy (cryo-EM) techniques. The first complete module from a bacterial modPKS, PikAIII in pikromycin biosynthesis, was reported in 2014 with back-to-back publications.[86, 87] In addition to the structure of a complete module (KS-AT-KR-ACP), the authors were able to selectively load individual domains with substrates to capture three different conformational states representing substrate loading, extension, and β-reduction that revealed substantial conformational movement of not only the ACP but also the AT and KR domains. Combined with an unusual orientation of the KS-AT portion, compared to mFAS or DEBS KS-AT structures, these findings were difficult to contextualize. Because there are so few structures of intact PKSs, it is unclear whether these unexpected conformations are common or rare among modPKS. As cryo-EM further matures, high-resolution structures of all classes of megasynthases will become more common, creating a library of PKS architectures in various states of programmed biosynthesis. The marriage of cryo-EM structures with site-specific cross-linking, high-resolution mass spectrometry, and improved heterologous expression systems promises that the next decade will bear exciting developments in our understanding of these complex nano-scale biological factories. The author hopes the contributions in this dissertation to the

understanding of programmed fungal *C*-methylation and editing will be built upon with these techniques.

1.5. References

1. Gilbert, J. A.; Jansson, J. K.; Knight, R., The Earth Microbiome project: successes and aspirations. *BMC Biol* **2014,** *12*, 69.
2. Drew, S. W.; Demain, A. L., Effect of primary metabolites on secondary metabolism. *Annu Rev Microbiol* **1977,** *31*, 343-56.
3. Vining, L. C., Secondary metabolism, inventive evolution and biochemical diversity--a review. *Gene* **1992,** *115* (1-2), 135-40.
4. Challis, G. L.; Hopwood, D. A., Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by Streptomyces species. *Proc Natl Acad Sci U S A* **2003,** *100 Suppl 2*, 14555-61.
5. Butler, M. S., Natural products to drugs: natural product derived compounds in clinical trials. *Nat Prod Rep* **2005,** *22* (2), 162-95.
6. Keller, N. P.; Turner, G.; Bennett, J. W., Fungal secondary metabolism - from biochemistry to genomics. *Nat Rev Microbiol* **2005,** *3* (12), 937-47.
7. Wong, F. T.; Khosla, C., Combinatorial biosynthesis of polyketides--a perspective. *Curr Opin Chem Biol* **2012,** *16* (1-2), 117-23.
8. Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* **2011,** *39* (Web Server issue), W339-46.
9. Walton, J. D., Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genet Biol* **2000,** *30* (3), 167-71.
10. Tianero, M. D.; Kwan, J. C.; Wyche, T. P.; Presson, A. P.; Koch, M.; Barrows, L. R.; Bugni, T. S.; Schmidt, E. W., Species specificity of symbiosis and secondary metabolism in ascidians. *ISME J* **2015,** *9* (3), 615-28.
11. Owen, J. G.; Charlop-Powers, Z.; Smith, A. G.; Ternei, M. A.; Calle, P. Y.; Reddy, B. V.; Montiel, D.; Brady, S. F., Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc Natl Acad Sci U S A* **2015,** *112* (14), 4221-6.
12. Lewis, K., Platforms for antibiotic discovery. *Nat Rev Drug Discov* **2013,** *12* (5), 371-87.
13. Chooi, Y. H.; Tang, Y., Navigating the fungal polyketide chemical space: from genes to molecules. *J Org Chem* **2012,** *77* (22), 9933-53.
14. Weissman, K. J.; Leadlay, P. F., Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol* **2005,** *3* (12), 925-36.

15. Poust, S.; Hagen, A.; Katz, L.; Keasling, J. D., Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform. *Curr Opin Biotechnol* **2014,** *30*, 32-9.
16. Ridley, C. P.; Lee, H. Y.; Khosla, C., Evolution of polyketide synthases in bacteria. *Proc Natl Acad Sci U S A* **2008,** *105* (12), 4595-600.
17. Fischbach, M. A.; Walsh, C. T., Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **2006,** *106* (8), 3468-96.
18. Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T., A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* **1996,** *3* (11), 923-36.
19. Wattana-amorn, P.; Williams, C.; Ploskon, E.; Cox, R. J.; Simpson, T. J.; Crosby, J.; Crump, M. P., Solution structure of an acyl carrier protein domain from a fungal type I polyketide synthase. *Biochemistry* **2010,** *49* (10), 2186-93.
20. Staunton, J.; Weissman, K. J., Polyketide biosynthesis: a millennium review. *Natural Product Reports* **2001,** *18* (4), 380-416.
21. Finking, R.; Marahiel, M. A., Biosynthesis of nonribosomal peptides1. *Annu Rev Microbiol* **2004,** *58*, 453-88.
22. Campbell, J. W.; Cronan, J. E., Jr., Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annu Rev Microbiol* **2001,** *55*, 305-32.
23. Leadlay, P. F., Combinatorial approaches to polyketide biosynthesis. *Curr Opin Chem Biol* **1997,** *1* (2), 162-8.
24. Kroken, S.; Glass, N. L.; Taylor, J. W.; Yoder, O. C.; Turgeon, B. G., Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc Natl Acad Sci U S A* **2003,** *100* (26), 15670-5.
25. Weissman, K. J., The structural biology of biosynthetic megaenzymes. *Nat Chem Biol* **2015,** *11* (9), 660-70.
26. Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L., Modular organization of genes required for complex polyketide biosynthesis. *Science* **1991,** *252* (5006), 675-9.
27. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G., Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc Natl Acad Sci U S A* **1999,** *96* (5), 1846-51.
28. McDaniel, R.; Kao, C. M.; Hwang, S. J.; Khosla, C., Engineered intermodular and intramodular polyketide synthase fusions. *Chem Biol* **1997,** *4* (9), 667-74.
29. Khosla, C., Harnessing the Biosynthetic Potential of Modular Polyketide Synthases. *Chem Rev* **1997,** *97* (7), 2577-2590.
30. Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C., Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem Biol* **1999,** *6* (2), 117-25.
31. Keatinge-Clay, A. T., The structures of type I polyketide synthases. *Nat Prod Rep* **2012,** *29* (10), 1050-73.
32. Helfrich, E. J.; Piel, J., Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* **2016,** *33* (2), 231-316.

33. O'Brien, R. V.; Davis, R. W.; Khosla, C.; Hillenmeyer, M. E., Computational identification and analysis of orphan assembly-line polyketide synthases. *J Antibiot (Tokyo)* **2014,** *67* (1), 89-97.
34. El-Sayed, A. K.; Hothersall, J.; Cooper, S. M.; Stephens, E.; Simpson, T. J.; Thomas, C. M., Characterization of the mupirocin biosynthesis gene cluster from Pseudomonas fluorescens NCIMB 10586. *Chem Biol* **2003,** *10* (5), 419-30.
35. Kusebauch, B.; Busch, B.; Scherlach, K.; Roth, M.; Hertweck, C., Functionally distinct modules operate two consecutive alpha,beta-->beta,gamma double-bond shifts in the rhizoxin polyketide assembly line. *Angew Chem Int Ed Engl* **2010,** *49* (8), 1460-4.
36. Bretschneider, T.; Heim, J. B.; Heine, D.; Winkler, R.; Busch, B.; Kusebauch, B.; Stehle, T.; Zocher, G.; Hertweck, C., Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **2013,** *502* (7469), 124-8.
37. Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A., Deconstruction of iterative multidomain polyketide synthase function. *Science* **2008,** *320* (5873), 243-6.
38. Ma, S. M.; Li, J. W.; Choi, J. W.; Zhou, H.; Lee, K. K.; Moorthie, V. A.; Xie, X.; Kealey, J. T.; Da Silva, N. A.; Vederas, J. C.; Tang, Y., Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **2009,** *326* (5952), 589-92.
39. Belecki, K.; Crawford, J. M.; Townsend, C. A., Production of Octaketide Polyenes by the Calicheamicin Polyketide Synthase CalE8: Implications for the Biosynthesis of Enediyne Core Structures. *J Am Chem Soc* **2009,** *131*, 12584-12566.
40. Belecki, K.; Townsend, C. A., Biochemical determination of enzyme-bound metabolites: preferential accumulation of a programmed octaketide on the enediyne polyketide synthase CalE8. *J Am Chem Soc* **2013,** *135* (38), 14339-48.
41. Rodel, T.; Gerlach, H., Enantioselective Synthesis of the Polyketide Antibiotic (3R,4S)-(-)-Citrinin. *Liebgs Ann* **1995**, 885-888.
42. Crawford, J. M.; Townsend, C. A., New insights into the formation of fungal aromatic polyketides. *Nat Rev Microbiol* **2010,** *8* (12), 879-89.
43. Minto, R. E.; Townsend, C. A., Enzymology and Molecular Biology of Aflatoxin Biosynthesis. *Chem Rev* **1997,** *97* (7), 2537-2556.
44. Shimizu, T.; Kinoshita, H.; Ishihara, S.; Sakai, K.; Nagai, S.; Nihira, T., Polyketide synthase gene responsible for citrinin biosynthesis in Monascus purpureus. *Appl Environ Microbiol* **2005,** *71* (7), 3453-7.
45. Daub, M. E.; Ehrenshaft, M., THE PHOTOACTIVATED CERCOSPORA TOXIN CERCOSPORIN: Contributions to Plant Disease and Fundamental Biology. *Annu Rev Phytopathol* **2000,** *38*, 461-490.
46. Daub, M. E.; Hangarter, R. P., Light-induced production of singlet oxygen and superoxide by the fungal toxin, cercosporin. *Plant Physiol* **1983,** *73* (3), 855-7.
47. Udwary, D. W.; Merski, M.; Townsend, C. A., A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *J Mol Biol* **2002,** *323* (3), 585-98.
48. Choquer, M.; Dekkers, K. L.; Chen, H.; Cao, L.; Ueng, P. P.; Daub, M. E.; Chung, K. R., The CTB1 Gene Encoding a Fungal Polyketide Synthase Is Required for Cercosporin Biosynthesis and Fungal Virulence of Cercospora nicotianae. *Mol Plant-Microbe Int* **2005,** *18* (5), 468-476.

49. Fujii, I.; Mori, Y.; Watanabe, A.; Kubo, Y.; Tsuji, G.; Ebizuka, Y., Heterologous expression and product identification of Colletotrichum lagenarium polyketide synthase encoded by the PKS1 gene involved in melanin biosynthesis. *Biosci Biotechnol Biochem* **1999,** *63* (8), 1445-52.
50. Ma, S. M.; Zhan, J.; Watanabe, K.; Xie, X.; Zhang, W.; Wang, C. C.; Tang, Y., Enzymatic Synthesis of Aromatic Polyketides Using PKS4 from Gibberella fujikuroi. *J Am Chem Soc* **2007,** *129*, 10642-10643.
51. Watanabe, A.; Fujii, I.; Sankawa, U.; Mayorga, M. E.; Timberlake, W. E.; Ebizuka, Y., Re-identification of Aspergillus nidulans wA Gene to Code for a Polyketide Synthase of Naphthopyrone. *Tetrahedron Letters* **1999,** *40*, 91-94.
52. Awakawa, T.; Yokota, K.; Funa, N.; Doi, F.; Mori, N.; Watanabe, H.; Horinouchi, S., Physically discrete beta-lactamase-type thioesterase catalyzes product release in atrochrysone synthesis by iterative type I polyketide synthase. *Chem Biol* **2009,** *16* (6), 613-23.
53. Vagstad, A. L.; Newman, A. G.; Storm, P. A.; Belecki, K.; Crawford, J. M.; Townsend, C. A., Combinatorial Domain Swaps Provide Insights into the Rules of Fungal Polyketide Synthase Programming and the Rational Synthesis of Non-Native Aromatic Products. *Angew Chem Int Ed Engl* **2013,** *52*, 1718-1721.
54. Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A., Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming. *J Am Chem Soc* **2014,** *136* (20), 7348-62.
55. Long, P. F.; Wilkinson, C. J.; Bisang, C. P.; Cortes, J.; Dunster, N.; Oliynyk, M.; McCormick, E.; McArthur, H.; Mendez, C.; Salas, J. A.; Staunton, J.; Leadlay, P. F., Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Mol Microbiol* **2002,** *43* (5), 1215-25.
56. Sun, H.; Ho, C. L.; Ding, F.; Soehano, I.; Liu, X. W.; Liang, Z. X., Synthesis of (R)-mellein by a partially reducing iterative polyketide synthase. *J Am Chem Soc* **2012,** *134* (29), 11924-7.
57. Kennedy, J.; Auclair, K.; Kendrew, S. G.; Park, C.; Vederas, J. C.; Hutchinson, C. R., Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* **1999,** *284* (5418), 1368-72.
58. Crawford, J. M.; Dancy, B. C.; Hill, E. A.; Udwary, D. W.; Townsend, C. A., Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proc Natl Acad Sci U S A* **2006,** *103* (45), 16728-33.
59. Cox, R. J.; Simpson, T. J., Fungal type I polyketide synthases. *Methods Enzymol* **2009,** *459*, 49-78.
60. Crawford, J. M.; Vagstad, A. L.; Whitworth, K. P.; Ehrlich, K. C.; Townsend, C. A., Synthetic strategy of nonreducing iterative polyketide synthases and the origin of the classical "starter-unit effect". *Chembiochem* **2008,** *9* (7), 1019-23.
61. Huitt-Roehl, C. R.; Hill, E. A.; Adams, M. M.; Vagstad, A. L.; Li, J. W.; Townsend, C. A., Starter unit flexibility for engineered product synthesis by the nonreducing polyketide synthase PksA. *ACS Chem Biol* **2015,** *10* (6), 1443-9.

62. Watanabe, C. M.; Townsend, C. A., Initial characterization of a type I fatty acid synthase and polyketide synthase multienzyme complex NorS in the biosynthesis of aflatoxin B(1). *Chem Biol* **2002,** *9* (9), 981-8.
63. Vagstad, A. L.; Bumpus, S. B.; Belecki, K.; Kelleher, N. L.; Townsend, C. A., Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *J Am Chem Soc* **2012,** *134* (15), 6865-77.
64. Fu, H.; Hopwood, D. A.; Khosla, C., Engineered biosynthesis of novel polyketides: evidence for temporal, but not regiospecific, control of cyclization of an aromatic polyketide precursor. *Chem Biol* **1994,** *1* (4), 205-10.
65. Crawford, J. M.; Korman, T. P.; Labonte, J. W.; Vagstad, A. L.; Hill, E. A.; Kamari-Bidkorpeh, O.; Tsai, S. C.; Townsend, C. A., Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **2009,** *461* (7267), 1139-43.
66. Xu, Y.; Zhou, T.; Zhou, Z.; Su, S.; Roberts, S. A.; Montfort, W. R.; Zeng, J.; Chen, M.; Zhang, W.; Lin, M.; Zhan, J.; Molnar, I., Rational reprogramming of fungal polyketide first-ring cyclization. *Proc Natl Acad Sci U S A* **2013,** *110* (14), 5398-403.
67. Korman, T. P.; Crawford, J. M.; Labonte, J. W.; Newman, A. G.; Wong, J.; Townsend, C. A.; Tsai, S. C., Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc Natl Acad Sci U S A* **2010,** *107* (14), 6246-51.
68. Newman, A. G.; Vagstad, A. L.; Belecki, K.; Scheerer, J. R.; Townsend, C. A., Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem Commun (Camb)* **2012,** *48* (96), 11772-4.
69. Vagstad, A. L.; Hill, E. A.; Labonte, J. W.; Townsend, C. A., Characterization of a fungal thioesterase having Claisen cyclase and deacetylase activities in melanin biosynthesis. *Chem Biol* **2012,** *19* (12), 1525-34.
70. Ahuja, M.; Chiang, Y. M.; Chang, S. L.; Praseuth, M. B.; Entwistle, R.; Sanchez, J. F.; Lo, H. C.; Yeh, H. H.; Oakley, B. R.; Wang, C. C., Illuminating the diversity of aromatic polyketide synthases in Aspergillus nidulans. *J Am Chem Soc* **2012,** *134* (19), 8212-21.
71. Ehmann, D. E.; Gehring, A. M.; Walsh, C. T., Lysine biosynthesis in Saccharomyces cerevisiae: mechanism of alpha-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. *Biochemistry* **1999,** *38* (19), 6171-7.
72. Gaitatzis, N.; Kunze, B.; Muller, R., In vitro reconstitution of the myxochelin biosynthetic machinery of Stigmatella aurantiaca Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A* **2001,** *98* (20), 11136-41.
73. Li, Y.; Weissman, K. J.; Muller, R., Myxochelin biosynthesis: direct evidence for two- and four-electron reduction of a carrier protein-bound thioester. *J Am Chem Soc* **2008,** *130* (24), 7554-5.
74. Davison, J.; al Fahad, A.; Cai, M.; Song, Z.; Yehia, S. Y.; Lazarus, C. M.; Bailey, A. M.; Simpson, T. J.; Cox, R. J., Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. *Proc Natl Acad Sci U S A* **2012,** *109* (20), 7642-7.

75. He, Y.; Cox, R. J., The molecular steps of citrinin biosynthesis in fungi. *Chem. Sci.* **2016,** *7* (3), 2119-2127.
76. Chen, W.; Chen, R.; Liu, Q.; He, Y.; He, K.; Ding, X.; Kang, L.; Guo, X.; Xie, N.; Zhou, Y.; Lu, Y.; Cox, R. J.; Molnar, I.; Li, M.; Shao, Y.; Chen, F., Orange, red, yellow: biosynthesis of azaphilone pigments in Monascus fungi. *Chem Sci* **2017,** *8* (7), 4917-4925.
77. Tang, Y.; Kim, C. Y.; Mathews, II; Cane, D. E.; Khosla, C., The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci U S A* **2006,** *103* (30), 11124-9.
78. Keatinge-Clay, A. T.; Stroud, R. M., The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure* **2006,** *14* (4), 737-48.
79. Zheng, J.; Taylor, C. A.; Piasecki, S. K.; Keatinge-Clay, A. T., Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase. *Structure* **2010,** *18* (8), 913-22.
80. Zheng, J.; Gay, D. C.; Demeler, B.; White, M. A.; Keatinge-Clay, A. T., Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat Chem Biol* **2012,** *8* (7), 615-21.
81. Barajas, J. F.; Shakya, G.; Moreno, G.; Rivera, H., Jr.; Jackson, D. R.; Topper, C. L.; Vagstad, A. L.; La Clair, J. J.; Townsend, C. A.; Burkart, M. D.; Tsai, S. C., Polyketide mimetics yield structural and mechanistic insights into product template domain function in nonreducing polyketide synthases. *Proc Natl Acad Sci U S A* **2017,** *114* (21), E4142-E4148.
82. Nguyen, C.; Haushalter, R. W.; Lee, D. J.; Markwick, P. R.; Bruegger, J.; Caldara-Festin, G.; Finzel, K.; Jackson, D. R.; Ishikawa, F.; O'Dowd, B.; McCammon, J. A.; Opella, S. J.; Tsai, S. C.; Burkart, M. D., Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **2014,** *505* (7483), 427-31.
83. Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T., Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proc Natl Acad Sci U S A* **2016,** *113* (7), 1802-7.
84. Maier, T.; Leibundgut, M.; Ban, N., The crystal structure of a mammalian fatty acid synthase. *Science* **2008,** *321* (5894), 1315-22.
85. Herbst, D. A.; Jakob, R. P.; Zahringer, F.; Maier, T., Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **2016,** *531* (7595), 533-7.
86. Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G., Structure of a modular polyketide synthase. *Nature* **2014,** *510* (7506), 512-7.
87. Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G., Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **2014,** *510* (7506), 560-4.

Chapter 2: Functional and Structural Analysis of Programmed *C*-Methylation in the Biosynthesis of the Fungal Polyketide Citrinin

This chapter was adapted with permission from P. A. Storm, D. A. Herbst, T. Maier, C. A. Townsend, *Cell Chem. Biol*. 2017, **24**, 316-325.

2.1. Introduction

PKSs and fatty acid synthases (FASs) utilize common strategies to produce an enormous number of natural products with highly diverse chemical scaffolds, including pigments, environmental toxins, and pharmacologically active substances.[1, 2] Simple precursor acyl substrates are tethered to the acyl carrier protein (ACP) by thioesterification of a post-translationally installed phosphopantetheine arm. They are then extended by the ketosynthase (KS) through a determined number of decarboxylative condensation reactions with ACP-bound malonyl extender units supplied by the malonyl-CoA:ACP transacylase (MAT). In many cases, the newly extended intermediate is delivered to additional tailoring domains that perform programmed modifications at the α- or β-carbon. Modular PKSs (modPKSs) generally have a distinct module for each round of extension and modification, whereas iterative PKSs (iPKSs) use the same domains repeatedly a defined number of times. How iPKS domains control which tailoring domains are used after a given round of extension is still largely unanswered and poses a significant challenge to engineering efforts.

Perhaps the simplest case of programmed tailoring among PKSs occurs during iterative *C*-methylation of fungal non-reduced polyketides, exemplified by the group VI and VII, or Clade III, families where a tetra- or pentaketide is *C*-methylated one or more times by an embedded *C*-methyltransferase (CMeT).[3, 4] Each round of extension produces

a potentially enolizable nucleophilic α-carbon, but only some such positions are methylated. A more complete understanding of how methylation is programmed and the structural basis by which methylation occurs will be essential for future efforts to rationally influence product methylation patterns. Site-selective polyketide *C*-methylation would be a powerful tool to engineer new biosynthetic pathways, and many methyltransferases can accept SAM analogs with larger alkyl donors, further expanding potential scaffolds.[5]

We selected PksCT as an initial candidate to explore non-reducing PKS (NR-PKS) *C*-methylation.[6] Responsible for the first isolable intermediate in citrinin (**1**) biosynthesis[7], PksCT produces a triply methylated pentaketide that is C2-C7 cyclized by the product template (PT) domain and released as the benzaldehyde (**2**) by the *C*-terminal reductase domain (R) (Figure 2.1).

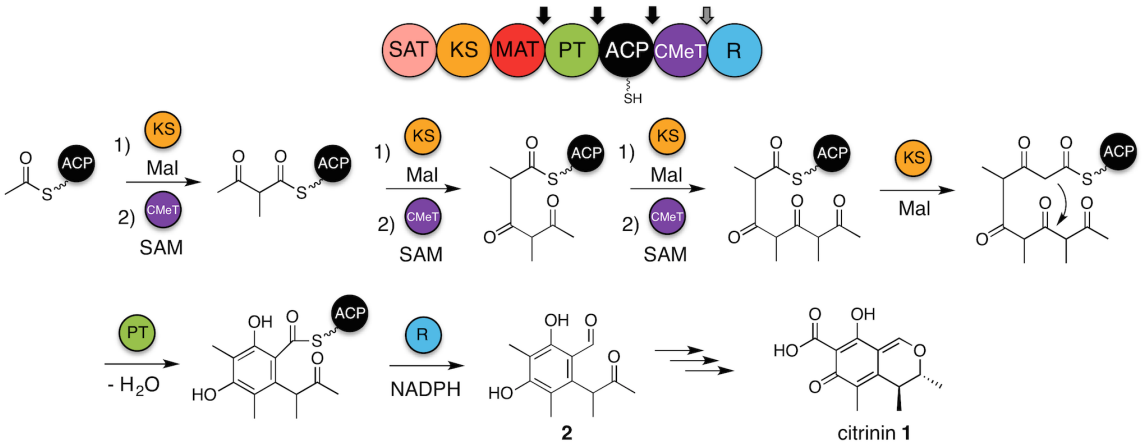


Figure 2.1: Proposed biosynthesis of 2 by PksCT. The domain architecture of PksCT is shown (top), with bold arrows indicating points of dissection (black: always used, gray: not used for CMeT-R didomain). Mal indicates a unit of ACP-bound malonyl used for decarboxylative condensation prior to methylation. Necessary domains and substrates are shown above and below the reaction arrows.

Previous efforts to understand NR-PKS programming have benefited greatly from a deconstruction approach that allows for the extension domains to be separated from the

tailoring domains.[8-10] A growing number of excised PKS domains have been structurally characterized, providing the molecular details of catalysis and substrate selection.[11] We sought to expand this experimental approach to PksCT in order to isolate CMeT domain activity and explore the structural basis for NR-PKS *C*-methylation programming.

2.2. Results

2.2.1. Domain Deconstruction and Reconstitution of PksCT

Using the Udwary-Merski algorithm (UMA)[12] to recognize interdomain linker regions in a set of related proteins, a domain deconstruction approach was applied to PksCT and used to generate mono- and multidomain fragments (Figure 2.2). A complete list of analyzed sequences and UMA parameters is available in Appendix A.

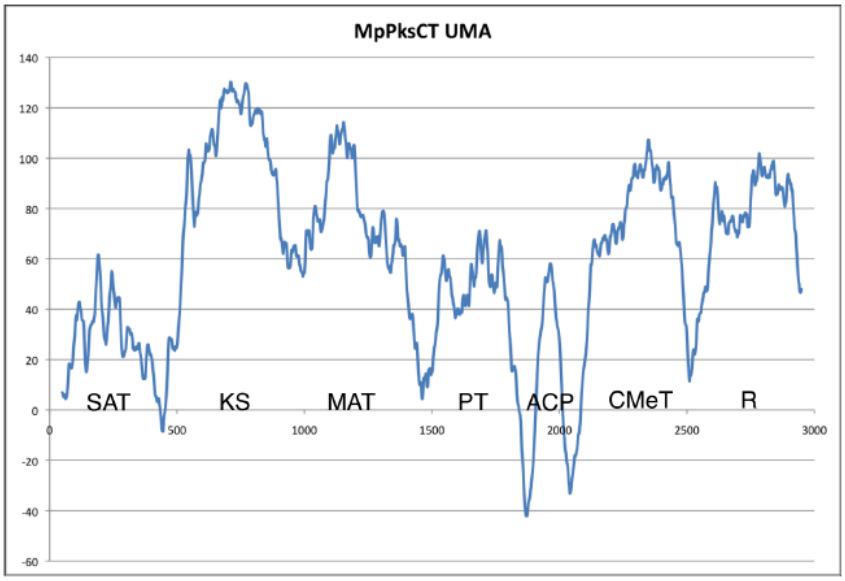


Figure 2.2: Domain deconstruction of PksCT using the Udwary-Merski algorithm. A curated list of NR-PKSs with conserved domain architecture was submitted to UMA, which scores each position of each entry based on sequence conservation, predicted secondary structure, and local hydrophobicity. Low scores indicate likely linker regions and were used to identify domain boundaries for preparation of mono- and multi-domain expression constructs.

All constructs gave soluble protein except for those containing the SAT domain in initial expressions. When multiple alternate domain boundaries did not improve expression, we reassessed the reported exons using the FGENESH exon prediction suite.[13] These results suggested alternate exon boundaries, which would include a conserved Trp195 in the SAT domain that was not present in the original annotated sequence. Total RNA was isolated from wild-type *Monascus purpureus* grown under citrinin producing conditions.[6, 14] The cDNA was synthesized, and a region containing the exon boundaries was amplified and subcloned. Five clones were sequenced and all verified that the revised prediction generated the correct exon boundaries (See Appendix A). We propose an update to GenBank: AB167465, joining gene fragments 1197-1835 and 1892-9034. PksCT residue numbering used hereafter reflects this revision.

Earlier reports speculated that PksCT could use several potential starter units[15], but heterologous *in vivo* extracts suggested an acetyl starter.[7] We confirmed this mode of initiation *in vitro* using a radiochemical assay.[16] PksCT SAT was capable of loading an acetyl starter unit from [1-^{14}C]-acetyl-CoA and transferring it to PksCT *holo*-ACP (See Appendix A). We then conducted a series of reconstitution reactions to identify the activity of the PksCT CMeT domain (Figure 2.3). Each individually expressed and purified fragment was combined as indicated with all of the expected substrates. Acetyl and malonyl units were included as the *N*-acetylcysteamine thioesters (SNAC), which serve as substitutes for the CoA thioesters used *in vivo*.[10] Stock solutions of SAM and NADPH were freshly prepared.

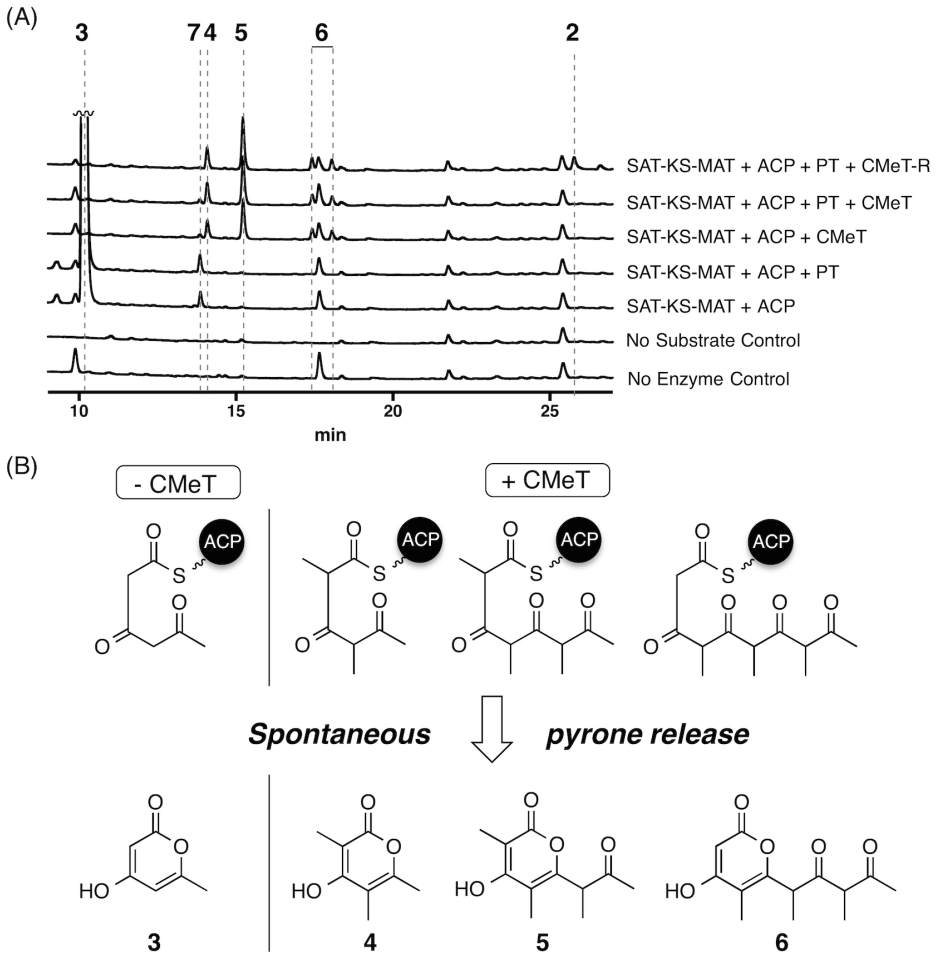


Figure 2.3: *In vitro* reconstitution of PksCT. (A) HPLC absorbance traces are shown at 280 nm for the combination of domains indicated to the right. Traces are vertically offset and peak 3 is truncated for clarity. (B) Spontaneous pyrone release of the intermediates as tri-, tetra-, and pentaketide pyrones allows for observation of the intermediates during each cycle. See Appendix A for UV-Vis and UPLC-ESI-MS data.

The minimal PKS, SAT–KS–MAT and *holo*-ACP, with acetyl and malonyl substrates primarily generated triketide **3**. Upon addition of PT, the product profile was not substantially different and no C2-C7 cyclized product was observed, even after attempts to chemically release any potentially cyclized intermediate with base or cysteamine.[17] Addition of CMeT and SAM to the minimal PKS, however, resulted in the loss of **3** but the appearance of four new species corresponding to spontaneous pyrone

release of methylated tri-, tetra-, and pentaketides **4**, **5**, and **6**. The pentaketide **6** eluted as two peaks, identical by mass, that appear to be diastereomers as each collected peak equilibrates to the same two components upon isolation and reinjection. The minimal PKS reactions also contained small amounts of **7**, having a mass consistent with a singly methylated form of **3**. The disappearance of **7** in reactions containing CMeT suggests that it results from an on-pathway intermediate to **4**, **5**, and **6**, and is discussed in detail below.

Full extension of the polyketide chain appears to rely upon the correct methylation of the growing linear intermediate, a feature seen previously in studies of the highly-reduced polyketide lovastatin.[18-20] This series of intermediates indicates that methylation occurs at the thioester α-carbon iteratively following each extension and not processively on a fully elongated pentaketide. These methylated derailment products are observed in the presence of the PT and R domains as well. With all domains, AcSNAC, MalSNAC, SAM, and NADPH present, the expected post-PKS aldehyde **2** was observed also, albeit in low yield.

Several efforts to improve the efficiency of reconstituted PksCT were attempted, including addition of methylthioadenosine nucleosidase (MTAN)[21] to hydrolyze the potentially inhibitory SAH co-product or inclusion of Mg^{2+}, a cofactor used by some methyltransferases.[22] Previous work has demonstrated that NR-PKS deconstruction reduces the yield of post-PKS product and increases the abundance of derailment products.[10, 23] In the case of deconstructed PksCT, it is possible that transfer of methylated polyketides from the KS active site to the CMeT active site allows for racemization of the methyl groups. Whether the KS, PT, or R domain active sites are structured for a specific methyl stereochemistry, and whether polyketides with scrambled

methyl stereochemistry are inhibitory, is currently unknown. Additionally, *in vivo* experiments of citrinin biosynthesis indicate that co-expression of an adjacent gene, encoding the hydrolase CitA, significantly improves titres of **2**, though the specific role of CitA is still unclear.[7]

2.2.2. Structure of PksCT CMeT in complex with SAH

The 1.65 Å crystal structure of the CMeT domain reveals an organization of two subdomains with the active site located at their interface (Figure 2.4). PksCT CMeT was co-crystallized with the stable co-product *S*-adenosylhomocysteine (SAH) in rhombohedral space group H3 containing one protomer per asymmetric unit (Appendix A), the structure was solved by SAD phasing and refined to R/R_{free} of 0.19/0.22.

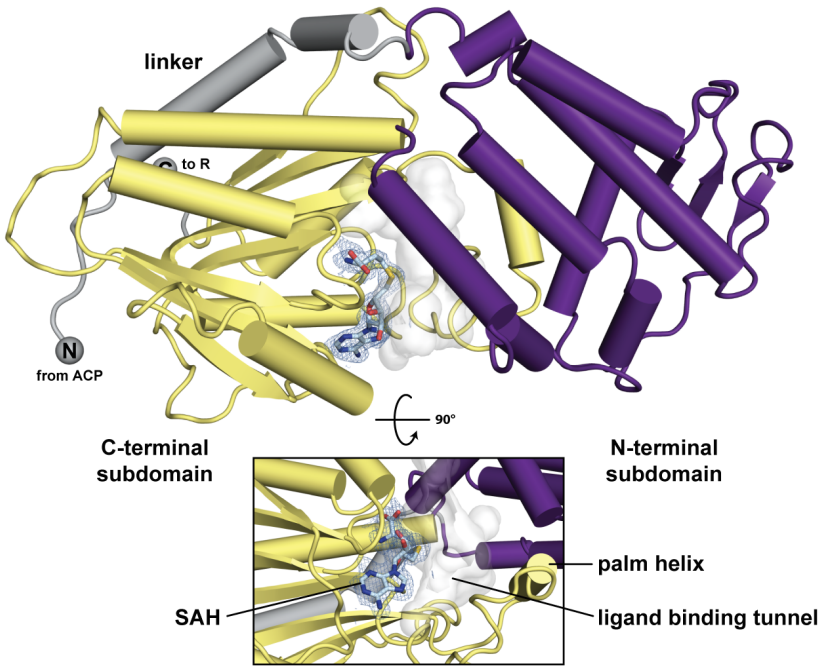


Figure 2.4: Crystal structure of PksCT CMeT in complex with SAH. PksCT CMeT is organized into an *N*-terminal linker (gray), and *N*-terminal subdomain (purple), and a *C*-terminal subdomain (yellow). The active site is located at the subdomain interface, and Fo – Fc omit difference density at 2.5σ level is shown for SAH.

The 202 amino acid (aa) *C*-terminal subdomain (residues 1959-2160) reveals a Class I methyltransferase fold, which is the most common methyltransferase fold in natural product biosynthetic enzymes, despite diverse acceptor substrates and low sequence conservation.[22, 24] The closest structural homologs are the SAM-dependent methyltransferase from *Aquifex aeolicus* (unpublished, PDB: 3DH0), the methyltransferase domain of bacterial-AvHen1-CN (PDB: 3JWJ)[25], a putative methyltransferase from *Sulfolobus solfataricus* (unpublished, PDB: 3I9F), NodS from *Bradyrhizobium japonicum* WM9 (PDB: 3OFK)[26], and the methyltransferase domain of bacterial-CtHen1-C (PDB: 3JWG)[25] with a C_α rmsd of 2.0-2.4 Å and 160-195 aligned residues. The catalytically inactive pseudo-methyltransferase (ΨCMeT) of the human fatty acid synthase (FAS) aligns with a C_α rmsd of 2.5 Å (158 aligned aa) and is the closest structural homolog in carrier protein-dependent multienzymes.[27] The *C*-terminal subdomain integrates into the *N*-terminal subdomain through a region around helix 15 that forms the base of the ligand binding tunnel, resembling an open palm ("palm helix region"), and is structurally conserved in mammalian FAS [27, 28], but deleted in insect FAS and some highly reducing-PKSs (HR-PKS).

The 144 aa *N*-terminal CMeT subdomain (residues 1815-1958) shares an interface of 1,515 $Å^2$ with the *C*-terminal subdomain and adopts an uncharacterized helical fold (Figure 2.5A). Such *N*-terminal subdomains of Class I methyltransferases that cap the SAH/SAM-binding pocket often serve a role in acceptor substrate selection.[29] In the related mammalian FAS multienzymes, which comprise catalytically inactive ΨCMeTs domains and provide the only structural depiction of CMeT integration into a multienzyme, this region is considerably truncated and disordered.[27, 28] Nevertheless, the

N-terminal CMeT subdomains in FASs reveal helices at equivalent positions to helix 4, 7, and 9/10 with a C_α rmsd of 2.5 Å over 42 aligned aa, indicating a common evolutionary origin (Figure 2.5B). The *N*-terminal subdomain is connected to the preceding protein regions in PksCT by a 30 aa long linker containing two helices and wraps around the *C*-terminal subdomain in a groove with an interface area of 1,461 $Å^2$. In mammalian FASs [27, 28], the *N*-terminal linker leading into ΨCMeT contacts this domain by appending a β-strand to the central β-sheet of the *C*-terminal subdomain (Figure 2.5B). In PksCT, however, the *N*-terminal linker (residues 1785-1814) wraps around the *C*-terminal subdomain in a prominent surface groove (Figure 2.5C).

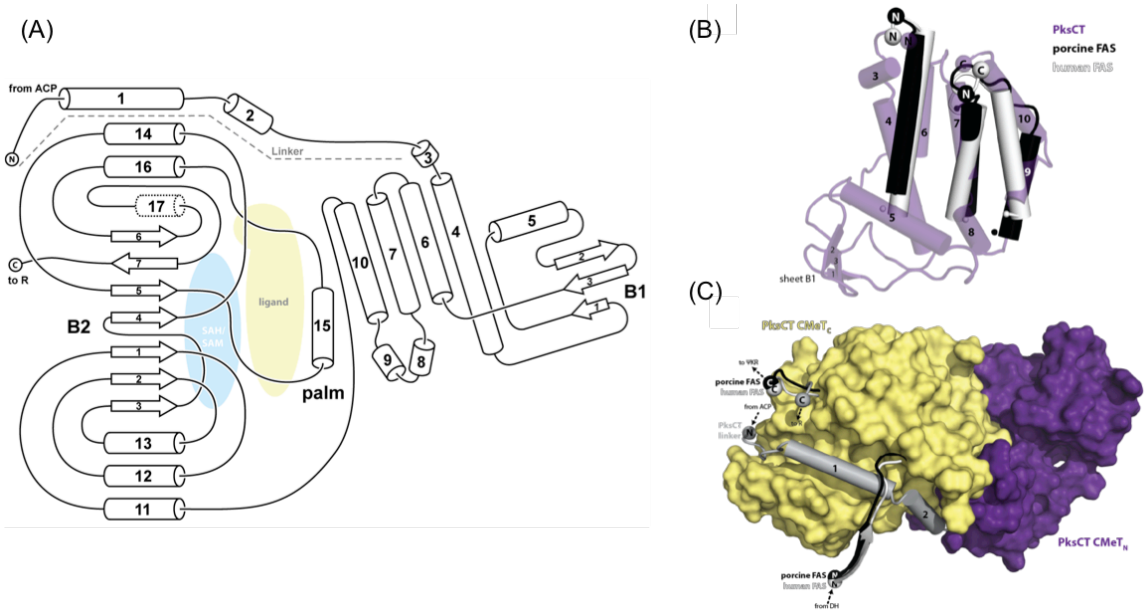


Figure 2.5: Topology of PksCT CMeT and comparison to FAS ΨCMeT. (A) PksCT CMeT topology revealed a previously uncharacterized N-terminal fold. α-helices are numbered, β-strands are numbered relative to their position in the respective β sheets B1 or B2, and helix 17 (dotted) is a 3_{10} helix. (B) A superposition of the truncated and largely disordered *N*-terminal subdomain of the animal FAS ΨCMeT (porcine: black; human: white) with PksCT CMeT indicates a structural conservation of three helices, overlapping with PksCT CMeT helices 4, 7, and 9/10. The ends of disordered regions are indicated by small spheres. Secondary structures are assigned according to Figure 2.5B. (C) In PksCT the *N*-terminal linker wraps in a surface groove around the *C*-terminal subdomain and positions the *N*- and *C*-terminal linker termini in close proximity to each other, whereas in FASs the *N*-terminal linker contacts this domain in a different location distant to the *C*-terminal linker terminus.

Consequently, in contrast to mammalian FAS, the *N*- and *C*-terminal linker ends are located in close proximity to each other, indicating an alternate integration of CMeT in the PksCT multienzyme.

The active site of PksCT CMeT is located at the interface of the *N*- and *C*-terminal subdomains. Most of the active site is formed by the *C*-terminal subdomain, which binds the SAH parallel to the ligand binding tunnel, which is lined with hydrophobic residues of the palm helix region. Helix 9 and 10 of the *N*-terminal subdomain contributes four residues to the distal region of the ligand binding tunnel to form the opposite face (Figure 2.6), and the highly conserved Tyr1955 splits the active site into a tunnel for the SAM co-substrate and the polyketide substrate. Remarkably, the homocysteine moiety of the SAH contributes to the formation of a pocket for extended substrates, which could accommodate varying polyketide chain lengths during elongation cycles 1-3.

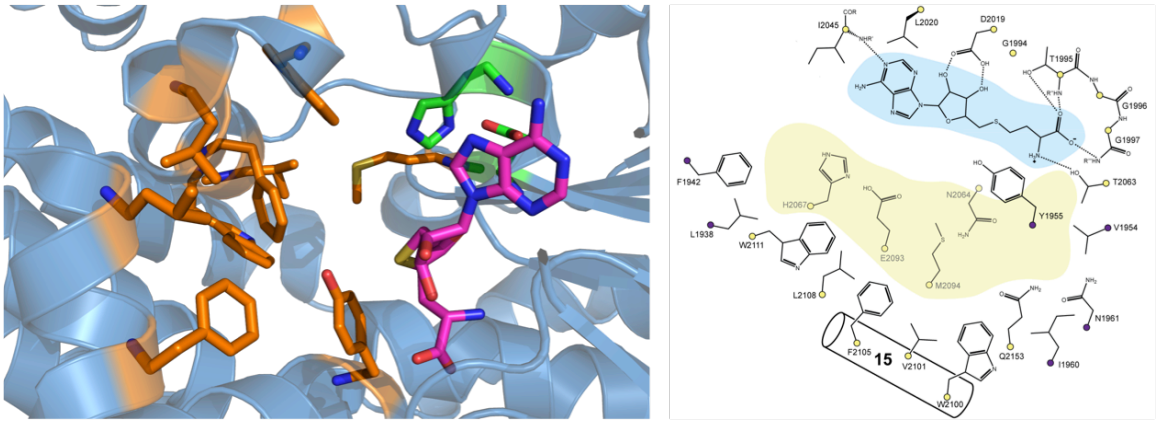


Figure 2.6: Active site of PksCT CMeT. (Left) The CMeT active site, as seen from the opening to solvent, is formed by the SAH cofactor (magenta) and several conserved residues, which are generally hydrophobic. The proposed base needed to deprotonate the α-carbon of the polyketide methyl-acceptor substrate is formed by the His-Glu dyad (green) positioned above the cavity. (Right) A flattened schematic of the active site, showing the conserved palm-helix residues. Blue and yellow shading indicates the SAM/SAH binding pocket and the acceptor substrate binding pocket respectively, as in Figure 2.5A.

The SAH is tightly bound at an interface of 356 $Å^2$ along one side of the co-substrate binding pocket via highly conserved residues: A GxGxGG (residues 1992-1997) motif forms hydrogen bonds with the homocysteine moiety, Asp2019 binds the ribose through two hydrogen bonds, and the adenine moiety is positioned in a hydrophobic pocket formed by Leu2020 and Ile2045 with a backbone hydrogen bond to Ile2045.

The active site tunnel reveals a funnel like shape and branches off into the extended substrate cavity at its inner end, providing a central channel for Cα substrate alignment with regard to different polyketide chain length. The central channel is surrounded by a patch of hydrophobic residues; the highly conserved Phe1942 as well as Leu1938 line the substrate entrance region and the conserved residues Leu2108 and Trp2111 and residues Trp2100, Val2101, Phe2105 of the palm helix would likely interact with the linear polyketide substrates. Together, these residues form a hydrophobic patch along which the hydrophobic backbone of substrates at various elongation states might slide into the active site. Interestingly, the residues involved in the hydrophobic patch are highly conserved among other NR-PKS CMeT domains – such as AusA, PkeA, MpaC, and DtbA – that install different methylation patterns than PksCT [3, 30, 31], and are functionally conserved in both *cis*- and *trans*-AT PKSs (Figure 2.7).

Although no obvious delineation among the NR-PKS CMeTs correlated with observed product methylation pattern, the number of available sequence–product pairs may be insufficient to identify specific programming motifs. Furthermore, a phylogenetic analysis of CMeT domains across multiple PKS/FAS families clearly demonstrated a branching between active CMeT and inactive ΨCMeTs. Active CMeTs

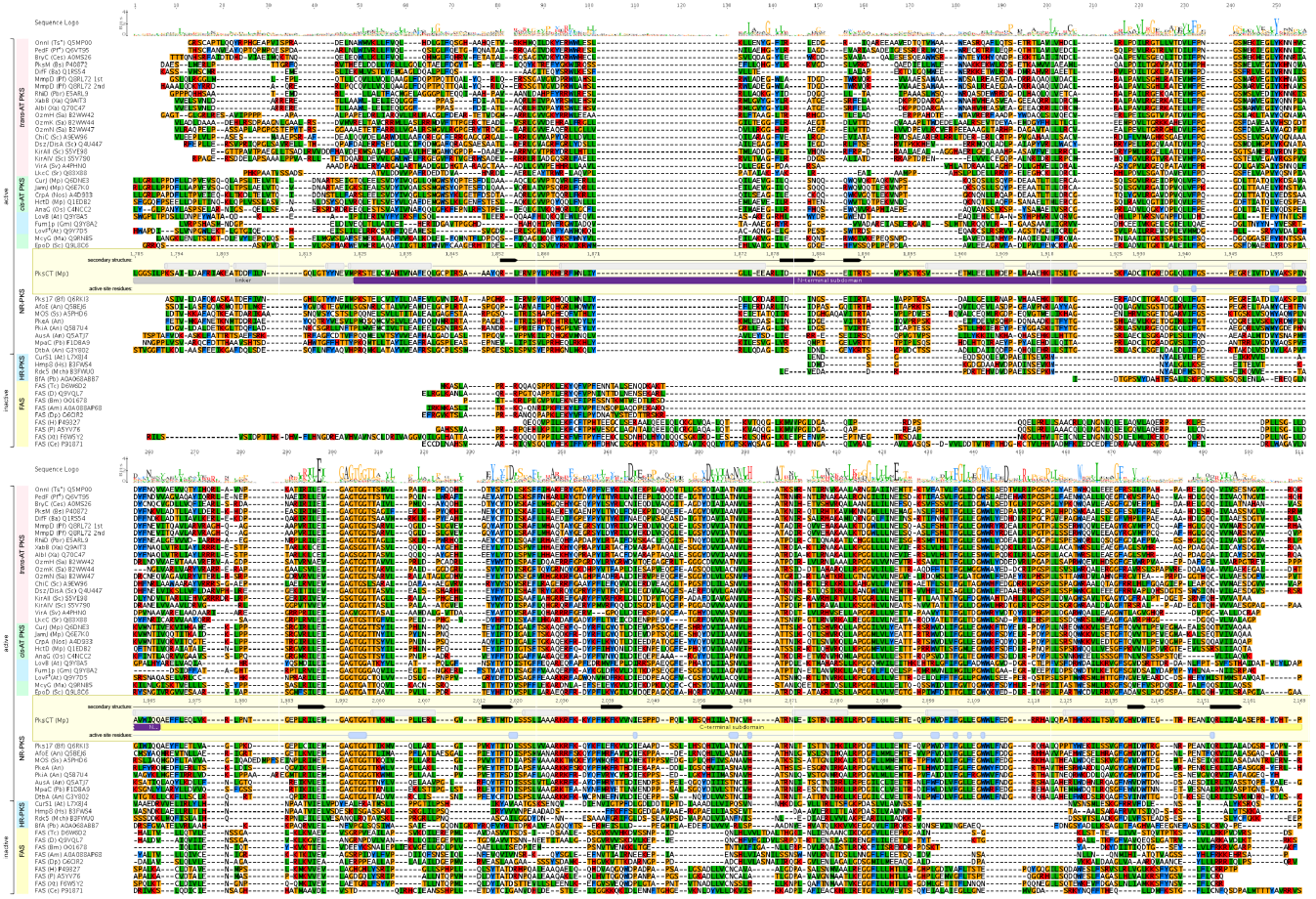


Figure 2.7: Alignment of 51 CMeT domains from PKSs and FASs. The alignment reveals conservation of active CMeT domains, whereas inactive ΨCMeT domains of FAS and some HR-PKS reveal deletions predominantly in the *N*-terminal subdomain and a low sequence conservation of active site residues.

clustered into three groups: *trans*-AT PKSs, *cis*-AT and HR-PKSs, and NR-PKSs (Figure 2.8). Active HR-PKS CMeT domains are nested among *cis*-AT CMeTs, perhaps reflecting similar embedding among reductive domains not present in NR-PKSs.

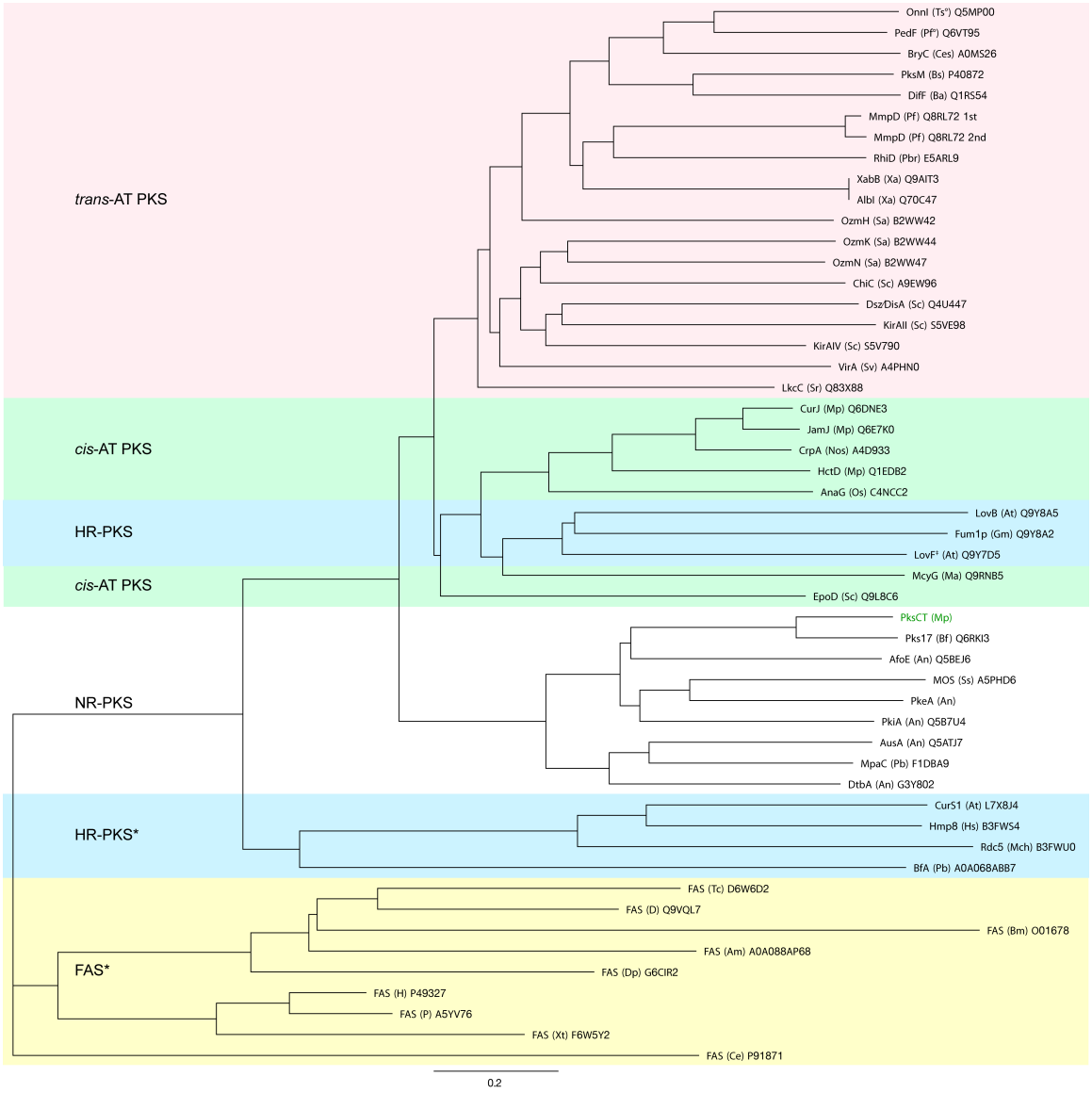


Figure 2.8: Phylogenetic analysis of 51 CMeT domains from PKSs and FASs. Thirty-eight active CMeT domains from PKSs and 13 inactive ΨCMeT from HR-PKSs and FASs were aligned (see Figure 2.7) and phylogenetically analyzed. Multienzyme family classifications are indicated in colored groups. Units are given as amino acid substitutions per site. All sequences are labeled as "protein name (organism abbrevation) UniProt number." The sequences of PksCT reflects the revised exons. (°: endosymbiont of listed organism, †: diketide synthase, *: inactive ΨCMeT domain).

2.2.3. Catalysis by CMeT

For catalysis, the α-carbon of the acceptor substrate presumably must be deprotonated to generate the enolate nucleophile for S_N2-like attack at the methyl donor. The invariant residues Tyr1955 and His2067 together with Glu2093 face the ligand binding tunnel from opposite sides and are well positioned to act as catalytic residues. His2067 forms a hydrogen bridge with Glu2093 and might act as catalytic dyad for enolization, as previously proposed for homocitrate synthase.[32]

Reconstitution of the minimal PksCT with the CMeT His2067Ala mutant primarily gave triketide **3**, matching the profile of the minimal PKS without CMeT (Figure 2.9).

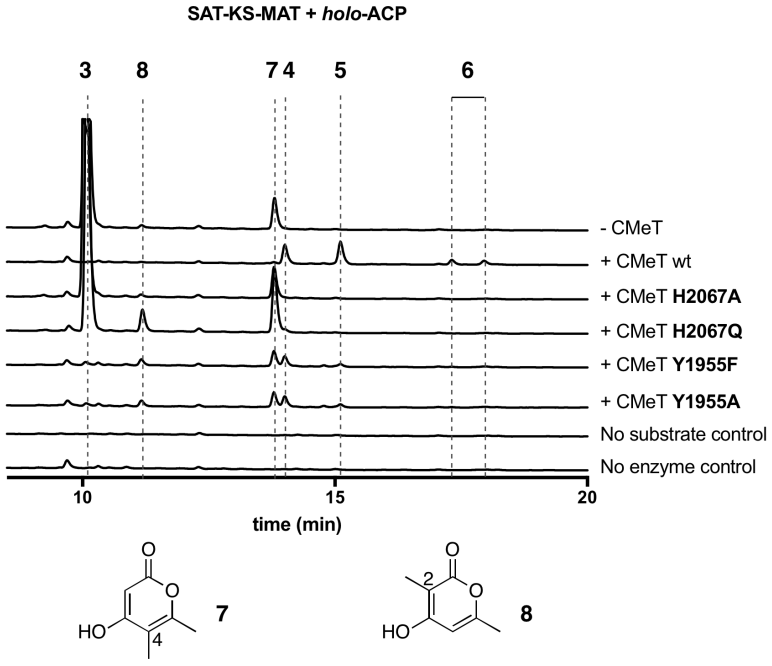


Figure 2.9: PksCT CMeT His2067 is essential for methyl transfer but Y1955 is less important. *In vitro* reconstitution of PksCT SAT-KS-MAT and *holo*-ACP with CMeT mutants show that methylated products **4**, **5**, and **6** are not produced by His2067 mutants. Tyr1955 mutants are capable of generating methylated triketides, but not pentaketides, suggesting a role for this residue in acceptor substrate binding. Absorbance traces are shown at 280 nm for the CMeT variant indicated on the right, and vertically offset with peak **3** truncated for clarity.

This experimental observation demonstrated the essential role of His2067 for efficient methyl transfer and supports its likely role as the catalytic base. The more conservative His2067Gln mutant[33, 34] also produced **3** as a major product in addition to smaller amounts of two different triketides **7** and **8** identical by mass, but singly methylated at C-4 and C-2, respectively. Tyr1955Phe and Tyr1955Ala mutations appeared to have a lesser effect on catalysis. Neither produced significant amounts of **3**, but instead yielded small amounts each of **4**, **5**, **7**, and **8**. The reduced amounts of **5** and the absence of **6**, compared to wild-type, suggest that Tyr1955 may be especially important for methylation of tetra- and pentaketide substrates.

Based on the experiments above, we propose a mechanism whereby the Glu-His dyad deprotonates the ACP-delivered acyl substrate at the α carbon of the thioester, forming a transient enolate nucleophile (Figure 2.10). If the orientation of the SAH co-product in the PksCT CMeT structure reflects a post-transfer conformation, the natural (*S*)-enantiomer of the sulfur center of SAM would point the methyl directly towards the interior of the binding pocket, well placed for nucleophilic attack. Following methyl transfer, both SAH and methylated acyl-ACP substrates must exit the active site, and the proton removed by the His-Glu dyad must exchange with solvent before the CMeT is catalytically competent. The order of substrate binding and release and the mechanism of proton release are not well understood, though the binding of the subsequent SAM, with the positively charged sulfonium, may influence egress of the proton.

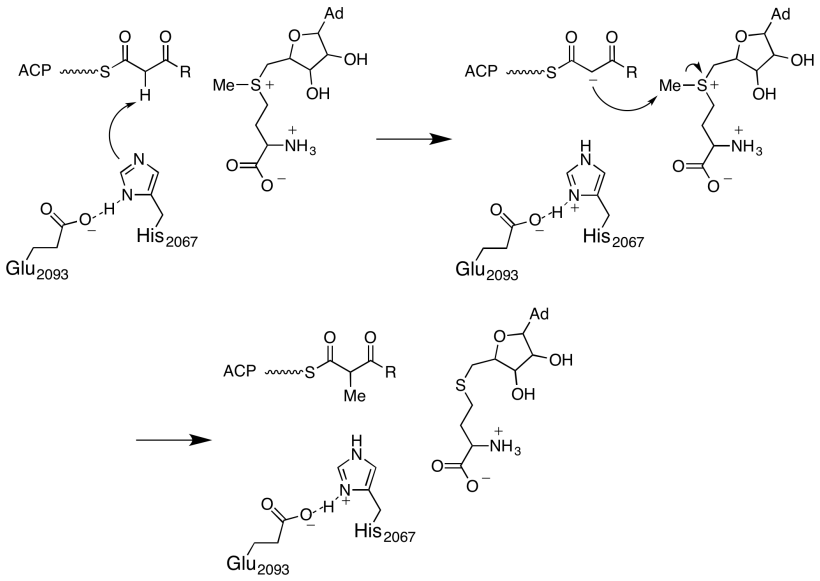


Figure 2.10: Proposed mechanism of methyl transfer. In the PksCT CMeT active site, His2067 forms a catalytic dyad with Glu2093 and deprotonates the α carbon to generate an enolate nucleophile capable of S_N2-like attack at the methyl donor. Completion of the catalytic cycle by loss of the removed proton to solvent is not explicitly shown. R indicates the potential chain lengths described in Figure 2.1.

Because **7** was observed also in the absence of CMeT, but appears to be on-path to **4**, **5**, and **6**, we speculate that some SAT–KS–MAT may be purified with a methylacetoacetyl unit bound to Cys555 in the KS active site. Such a substrate could arise from direct loading from methylacetoacetyl-CoA, an intermediate in isoleucine catabolism[35] or perhaps homologation of KS-bound acetyl by endogenous methylmalonyl-CoA during heterologous expression. (For an example of closely related biochemistry, see Steyn and Vleggaar, 1984 and Steyn et al., 1981.)[36, 37] A single round of extension of methylacetoacetyl, followed by spontaneous pyrone formation would result in **7**, methylated at C-4. The decreased activity of the CMeT His2067Gln mutant (Figure 2.9) could allow for a relative increase in **7** and the appearance of **8**, methylated at C-2, where the first potential methylation is skipped but the second methylation is successful. The identity of 8 was subsequently confirmed by comparison to an authentic standard, verifying the methyl position. It is known that polyketide extension is

exceedingly rapid in NR-PKSs.[8, 38] Kinetic competition between chain elongation and programmed methylation where the latter is impaired in the His2067Gln mutant could lead to expected alkylation at C-4 to be missed.

2.3. Discussion

Methylation is a common and powerful strategy to diversify the structures of microbial natural products. iPKSs incorporate *C*-methyl groups into their product scaffolds early in the biosynthetic pathway because the pK_a of the methylated position is generally never lower or more accessible to enzymatic chemistry than immediately following extension to the β-ketothioester, particularly for PKSs that reductively tailor the β position. Early indications of the interplay between KS and CMeT domains were observed in LovB, where the lack of SAM or CMeT resulted in the accumulation of truncated intermediates.[18, 20] Recently, it was shown that the LovB CMeT domain outcompetes the LovB KR activity by virtue of its higher catalytic efficiency.[18] Non-cognate substrates, due to incorrect modifications in previous cycles, caused significant attenuation of activity for the highly specific CMeT, resulting in off-loading of those substrates. This finding suggested a role of CMeT as gatekeeper, which has to be passed after three of six cycles of polyketide chain extension. Unlike for LovB CMeT, PksCT CMeT is not competing with a KR domain and methylates multiple substrates. Therefore, PksCT CMeT appears to have a similar function by adding methyl groups as check-point tags, which are recognized by PksCT KS, such that a lack of methylation causes release of immature products at the triketide stage. Abortive release of an off-pathway

intermediate may be useful for NR-PKSs that lack a *C*-terminal thioesterase domain to serve an editing role.[38]

Any future efforts to modulate the pattern of *C*-methylation in iPKSs may require perturbation of both CMeT and KS substrate binding sites. The highly conserved residues of the acceptor binding pocket also indicate that methylation programming is determined by other features of the active site or acceptor:CMeT complex. That programmed methylation is retained even after domain dissection, however, points towards control residing in intrinsic features of the CMeT. Additional studies will be needed to identify and manipulate these features to control programmed methylation. The existence of *gem*-dimethylating CMeTs in *trans*-AT PKSs[39] combined with the iterative action of NR-PKS CMeTs show that considerable flexibility is possible, potentially allowing for engineered products.

Sequence motifs have been identified in other iPKS modification domains that allow for prediction of the stereochemical outcome, notably the A- and B-type ketoreductase (KR) domains.[11] A similar dichotomy for CMeTs, pro-*R* or pro-*S*, has not been identified. The stereochemistry of methylation could not be inferred from our data, and in the context of NR-PKSs this information is lost upon aromatization. Future structural work that includes an acceptor substrate should provide greater detail about the methyl transfer mechanism.

In connection with these findings, a structure of an excised CMeT from a bacterial modular PKS was determined contemporaneously and closely agrees with our findings.[40] Briefly, the CMeT domain of CurJ, a module used in curacin biosynthesis, was found to install a single methyl group on an elongated, highly-reduced substrate. While some

bacterical *C*-methyl groups are installed by methylating malonyl units prior to extension[41], CurJ was shown to methylate the post-extension ACP-bound β-ketothioester like PksCT CMeT, and had no activity towards malonyl substrate mimics. A structural comparison of CurJ CMeT and PksCT CMeT reveals an identical subdomain organization, including the helical *N*-terminal subdomain, though CurJ crystallized as a dimer instead of a monomer (PDB: 5THZ). A pairwise structural alignment of PksCT CMeT to both protomers of CurJ CMeT returned 1.9 and 2.0 Å over 315 and 321 aligned residues.[42] The authors were unable to capture CurJ in complex with the polyketide substrate, but the active site of CurJ also includes the conserved His-Glu dyad and generally hydrophobic residues lining a deep pocket in the cleft between the two subdomains. The strong structural similarities between these two CMeT domains, despite their disparate sources and substrates, suggest that intrinsic but unknown features of the CMeT are responsible for modulating their activity. The effects of the acyl-ACP binding to the CMeT are not known, but an attractive area for future study.

2.4. Conclusions

Structure determination of the PksCT CMeT domain revealed a characteristic organization of the *N*-terminal subdomain in active PKS CMeT domains and a detailed description of residues lining the substrate binding site. Structural alignments to mammalian FAS and sequence comparison to modular *cis*- and *trans*-AT PKS also demonstrate a conserved organization of inactive ΨCMeT domains, commonly observed in HR-PKS and FAS and active CMeT domains in all systems. The PksCT CMeT structure thus provides a blueprint for the systematic analysis and variation of CMeT

functions across all PKS systems. Despite a high degree of conservation at the level of subdomain organization and active site assembly, the organization of *N*-terminal linkers in the CMeT domain of the NR-PKS PksCT clearly deviates from those observed for the ΨCMeT domain of fully-reducing mammalian FAS, indicating distinct modes of integration in the two systems. The exact modes of domain integration (and the extent of their conservation) of CMeT domains in systems other than FAS, namely iterative and modular *cis*- and *trans*-AT PKS, remains to be explored. First indications may be obtained from a careful comparison of flanking regions in excised CMeT domains. Differences in intrinsic methylation patterns may also be explained through multi-domain structures and mutagenesis of binding pocket residues.

2.5. Experimental

2.5.1. Materials

Reagents were purchased from Sigma Aldrich (St. Louis, MO) unless stated otherwise. Standard molecular biology procedures were used for nucleic acid isolation and assembly of expression constructs. Sequencing was done by the Johns Hopkins University Synthesis and Sequencing Facility (Baltimore, MD). *M. purpureus* NRRL 1596 was received from the ARS, USDA (Peoria, IL) and cultured on PDA plates. Genomic DNA was extracted from mycelia using the DNeasy Plant Mini kit (Qiagen, Hilden, Germany). Target sequences were amplified by polymerase chain reaction using the primers listed in Appendix A.

2.5.2. Sequence analysis, cloning, and protein production

To verify the mRNA sequence of *pksCT*, *M. purpureus* NRRL 1596 was grown in MC medium following reported methods for citrinin production.[6] After 7 days of growth at 28 °C, the mycelia were harvested by centrifugation and lyophilized. The mycelial powder was ground under liquid nitrogen and total RNA was purified using the RNeasy Plant Mini kit (Qiagen, Hilden, Germany) following the manufacturer's protocol with on-column DNA digestion. A cDNA library was prepared using 1 μg total RNA with 0.5 μg random hexamer (Applied Biosystems, Foster City, CA) and M-MLV reverse transcriptase (Promega, Madison, WI) at 37 °C for 1 h. Negative control reactions lacked any reverse transcriptase. PCR amplification of the desired exon boundaries was done using the cDNA library products directly (2 μL) with primers MpPksCTex1-5 and MpPksCT-SAT-3 and Phusion polymerase (New England Biolabs, Ipswich, MA). Primers for the constitutively expressed actin gene MpAct5 and MpAct3, based on a previous report[6], were used as positive controls. The PCR reactions were separated by 1% agarose gel and the lone product band was ligated into pCR-blunt (Life Technologies, Carlsbad, CA) and transformed into DH5α by heat shock treatment. Five clones were sequenced using M13F and M13R primers.

Domain boundaries and expression constructs are detailed in Appendix A. Each construct was transformed to *E. coli* BL21(DE3) and grown in LB media at 37 °C to OD_{600} of 0.6. The cultures were cooled in an ice bath for 30-60 min, and expression was induced with 1 mM IPTG (GoldBio, St. Louis, MO) overnight at 18 °C. Cell pellets were harvested by centrifugation for 15 min, at 4,000 x g, 4 °C. Pellets were either flash frozen in $N_2(l)$ and stored at -80 °C until use, or resuspended immediately in 5 mL/g cell pellet

in buffer A (50 mM potassium phosphate, 300 mM NaCl, 10 % glycerol, pH 7.6). The slurry was sonicated on ice for 10 x 10 s at 40% amplitude (Vibra-Cell Ultrasonic Processor, Sonics & Materials, Inc., Newtown, CT) and the lysate was cleared by centrifugation for 25 min at 27,000 x g, 4 °C. The cleared lysate was batch bound to Co^{2+}-TALON (Clontech, Mountain View, CA) at 4 °C for 1 h. Purification was performed by gravity column and the applied resin was washed with buffer A (10 CV) and buffer A + 2 mM imidazole (5 CV) before eluting in buffer A + 100 mM imidazole (5 CV). Each domain fragment was dialyzed into reaction buffer (100 mM potassium phosphate, 10 % glycerol, pH 7.0) and used immediately or flash frozen in $N_2(l)$ and stored at -80 °C until use.

For crystallization, PksCT CMeT was expressed and initially purified as described above, but dialyzed into 25 mM Tris, 5 % glycerol, pH 7.5 (buffer B). A secondary purification was performed by applying 60 mg PksCT CMeT to 25 mL Q-Sepharose Fast-flow resin pre-equilibrated in buffer B. A series of 2 CV step-wise washes of buffer B + 0, 10, 50, 100, 150, 200, 300, and 500 mM KCl were then applied to the column, with the CMeT eluting in the 50 and 100 mM KCl fractions. These fractions were combined, dialyzed into 2 x 1 L buffer B, and concentrated to 10 mg/mL.

Selenomethionine-labeled (SeMet) CMeT was prepared by growth in M9 minimal media to OD_{600} = 0.6, at which point the media was supplemented with the following L-amino acids per 1 L: 100 mg each lysine, phenylalanine, and threonine; 50 mg each isoleucine, leucine, and valine; 30 mg selenomethionine.[43] The cultures were allowed to continue to grow at 37 °C for 15 min before being cooled and inoculated as decribed

above. Purification of SeMet CMeT for crystallization was identical to that of the native CMeT.

2.5.3. In vitro reconstitution of PksCT and analysis of product formation

Protein concentration was determined by Bradford assay (BioRad, Hercules, CA) in duplicate using bovine serum albumin as a standard. Prior to *in vitro* reactions, ACP was activated by Sfp with CoASH and $MgCl_2$ in reaction buffer for 1 h at room temperature as previously reported.[10] PksCT reconstitution reactions contained 10 μM of each included domain with 0.5 mM AcSNAC, 2 mM MalSNAC, 2 mM SAM, 1 mM NADPH, and 1 mM TCEP in reaction buffer totaling 250 μL. AcSNAC was prepared synthetically and MalSNAC was purified from MatB reactions.[38] After 4 h at room temperature, the reactions were quenched with 5 μL concentrated HCl and extracted into ethyl acetate 3 x 250 μL. The combined organic extracts were dried to a residue and resuspended in 250 μL of 20 % aqueous acetonitrile. Extracts were analyzed on an Agilent 1200 HPLC with autosampler by injecting 100 μL onto a Prodigy ODS3 column (4.5 x 250 mm, 5μ, Phenomenex, Torrence, CA) with 5-85 % MeCN/H_2O, with 0.1 % formic acid, over 40 min at 1 mL/min. Mass spectrometric analysis was done using a Waters Acquity Xevo G-2 UPLC-ESI-MS in positive ion mode, with 5 μL injected on a BEHC C18 column and 10-90 % MeCN/H_2O, with 0.1 % formic acid, over 10 min.

Radiolabel transfer assays were carried out with 10 μM each protein in the above reaction buffer. Reactions were initiated with the addition of [1-^{14}C]-acetyl-CoA (American Radiolabeled Chemicals, St. Louis, MO) to 50 μM at room temperature for 5 min and quenched with the addition of 5X SDS loading buffer. The samples were

separated by 16 % SDS-PAGE and dried. The dried gel was exposed to a phosphorimager screen (Amersham, Piscataway, NJ). Data was collected on a Typhoon 9410 Variable Mode imager (Amersham, Piscataway, NJ) and analyzed using ImageJ.[44]

2.5.4. CMeT crystallization

Protein stock solutions were dialyzed (4x, 1:2000) against crystallization buffer (25 mM Tris-HCl, 5 % (v/v) glycerol, 2 mM DTT, pH 7.0). All crystallization experiments were performed using a robotic setup applying the sitting drop vapor diffusion method. Crystals in space group H3 were grown at 17 °C by mixing 0.5 µL of protein at 14.9 mg/mL (supplemented with 2 mM SAH) in crystallization buffer with 0.5 µL reservoir solution (0.1 M BIS-TRIS pH 6.57, 25% (w/v) PEG MME 2K) and grew to a final size of 0.4x0.2x0.1 mm^3 within three days. Prior to harvesting, crystals were cryo-protected by slowly exchanging the drop solution (0.1 M BIS-TRIS 8.0 pH, 25 % (w/v) PEG MME 2K, 10 mM acetoacetyl-CoA, 20 % (v/v) ethylene glycol) followed by an incubation of 5 min and flash freezing in $N_2(l)$. SeMet CMeT was crystallized in space group H3 at 17 °C by mixing 0.5 µL of protein at 14.3 mg/mL (supplemented with 2 mM SAH) in crystallization buffer with 0.5 µL reservoir solution (0.1 M BIS-TRIS pH 6.21, 21.4% (w/v) PEG MME 2K). Crystals were harvested after four days at a size of 0.4x0.3x0.05 mm^3. Prior to harvesting and flash freezing in $N_2(l)$, crystals were dehydrated and cryo-protected (0.1 M BIS-TRIS pH 6.5, 30.0 % (w/v) PEG MME 2K, 2 mM SAH, 25 % (v/v) ethylene glycol) by carefully exchanging the drop solution over a period of 30-60 min.

2.5.5. Data collection, structure determination and analysis

Data sets were collected at the Swiss Light Source (SLS, Villigen, Switzerland) at beamline X06DA (native: λ = 0.99998 Å and 1.90747 Å; SeMet: λ = 0.97929 Å) and a temperature of 100 K. Data reduction was performed using DIALS[45] and XDS[46], datasets were analyzed with Phenix[47]. Resolution cutoffs were determined by $CC_{1/2}$ criterion.[48] The structure of SeMet CMeT was solved by SAD phasing using SHELX.[49] The asymmetric unit contained one protomer, which was used for initial phasing of native crystals diffracting to higher resolution using molecular replacement in PHASER.[50] Initial maps were improved by density modification using PARROT[51], followed by automated rebuilding with BUCCANEER.[52] The final model was obtained after iterative cycles of model building in COOT[53] and TLS refinement in Phenix[47], yielding excellent geometry (Ramachandran favored/outliers: 99.48 % / 0.00 %) and R_{work}/R_{free} values of 0.19 / 0.22 (Appendix A). The final model covers residues 1785-2163. All structures of native and SeMet-labeled protein revealed weak, but significant Fo-Fc difference density for a partially occupied ligand in the ligand binding tunnel (Appendix A), which could not be displaced by acetoacetyl-CoA soaks (cryo solutions + 15 mM acetoacetyl-CoA). Anomalous data collection at λ = 1.90747 Å did not reveal significant anomalous signal as expected for a fully occupied sulfur or phosphorus atom as part of the ligand (Appendix A). The observation was verified by two protein purifications and crystal treatments in the presence or absence of acetoacetyl-CoA. Despite the featured shape of the density, the nature and origin of the ligand, either from the crystallization conditions or protein purification, could not be unambiguously determined.

Related folds were identified using the protein structure comparison service PDBeFold at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm), authored by E. Krissinel and K. Henrick[42], using 40 % (query) and 70 % (target) as matching parameters, and interfaces were analyzed using QtPISA.[54] Bias-removal for Fo-Fc omit maps was achieved by applying a random perturbation to coordinates (Δ0.2 Å) and B-factors (Δ20 % of the mean overall B-factor) using MOLEMAN2[55] prior to refinement. Figures were generated using PYMOL.[56]

2.5.6. Phylogenetic sequence analysis

Fifty-one sequences of CMeT domains from NR-, HR-, *cis*-AT- and *trans*-AT PKS as well as FAS were selected to analyze phylogenetic distances. Sequences of inactive pseudo CMeT domains were included for FAS and HR-PKS. Alignments were generated using Clustal Omega[57] and phylogenetic trees were generated using the neighbor joining algorithm implemented in Geneious version 8.1.6 (http://www.geneious.com).[58]

2.5.7 Accession Number

The atomic coordinates and structure factors have been deposited in the PDB (http://wwpdb.org) under the accession code PDB: 5MPT.

2.6. References

1. Weissman, K. J.; Leadlay, P. F., Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **2005,** *3* (12), 925-36.
2. Crawford, J. M.; Townsend, C. A., New insights into the formation of fungal aromatic polyketides. *Nat. Rev. Microbiol.* **2010,** *8* (12), 879-89.
3. Ahuja, M.; Chiang, Y. M.; Chang, S. L.; Praseuth, M. B.; Entwistle, R.; Sanchez, J. F.; Lo, H. C.; Yeh, H. H.; Oakley, B. R.; Wang, C. C., Illuminating the diversity of aromatic polyketide synthases in Aspergillus nidulans. *J. Am. Chem. Soc.* **2012,** *134* (19), 8212-21.
4. Kroken, S.; Glass, N. L.; Taylor, J. W.; Yoder, O. C.; Turgeon, B. G., Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc Natl Acad Sci U S A* **2003,** *100* (26), 15670-5.
5. Struck, A. W.; Thompson, M. L.; Wong, L. S.; Micklefield, J., S-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *ChemBioChem* **2012,** *13* (18), 2642-55.
6. Shimizu, T.; Kinoshita, H.; Ishihara, S.; Sakai, K.; Nagai, S.; Nihira, T., Polyketide synthase gene responsible for citrinin biosynthesis in Monascus purpureus. *Appl Environ Microbiol* **2005,** *71* (7), 3453-7.
7. He, Y.; Cox, R. J., The molecular steps of citrinin biosynthesis in fungi. *Chem. Sci.* **2016,** *7* (3), 2119-2127.
8. Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A., Deconstruction of iterative multidomain polyketide synthase function. *Science* **2008,** *320* (5873), 243-6.
9. Vagstad, A. L.; Newman, A. G.; Storm, P. A.; Belecki, K.; Crawford, J. M.; Townsend, C. A., Combinatorial Domain Swaps Provide Insights into the Rules of Fungal Polyketide Synthase Programming and the Rational Synthesis of Non-Native Aromatic Products. *Angew. Chem. Int. Ed. Engl.* **2013,** *52*, 1718-1721.
10. Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A., Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming. *J. Am. Chem. Soc.* **2014,** *136* (20), 7348-62.
11. Keatinge-Clay, A. T., The structures of type I polyketide synthases. *Nat. Prod. Rep.* **2012,** *29* (10), 1050-73.
12. Udwary, D. W.; Merski, M.; Townsend, C. A., A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *J. Mol. Biol.* **2002,** *323* (3), 585-98.
13. Solovyev, V.; Kosarev, P.; Seledsov, I.; Vorobyev, D., Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol* **2006,** *7 Suppl 1*, S10 1-12.
14. Sakai, K.; Kinoshita, H.; Shimizu, T.; Nihira, T., Construction of a citrinin gene cluster expression system in heterologous Aspergillus oryzae. *J. Biosci. Bioeng.* **2008,** *106* (5), 466-72.

15. Davison, J.; al Fahad, A.; Cai, M.; Song, Z.; Yehia, S. Y.; Lazarus, C. M.; Bailey, A. M.; Simpson, T. J.; Cox, R. J., Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. *Proc Natl Acad Sci U S A* **2012,** *109* (20), 7642-7.
16. Foulke-Abel, J.; Townsend, C. A., Demonstration of starter unit interprotein transfer from a fatty acid synthase to a multidomain, nonreducing polyketide synthase. *ChemBioChem* **2012,** *13* (13), 1880-4.
17. Belecki, K.; Townsend, C. A., Biochemical determination of enzyme-bound metabolites: preferential accumulation of a programmed octaketide on the enediyne polyketide synthase CalE8. *J. Am. Chem. Soc.* **2013,** *135* (38), 14339-48.
18. Cacho, R. A.; Thuss, J.; Xu, W.; Sanichar, R.; Gao, Z.; Nguyen, A.; Vederas, J. C.; Tang, Y., Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase. *J. Am. Chem. Soc.* **2015,** *137* (50), 15688-91.
19. Fisch, K. M.; Bakeer, W.; Yakasai, A. A.; Song, Z.; Pedrick, J.; Wasil, Z.; Bailey, A. M.; Lazarus, C. M.; Simpson, T. J.; Cox, R. J., Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite. *J. Am. Chem. Soc.* **2011,** *133* (41), 16635-41.
20. Ma, S. M.; Li, J. W.; Choi, J. W.; Zhou, H.; Lee, K. K.; Moorthie, V. A.; Xie, X.; Kealey, J. T.; Da Silva, N. A.; Vederas, J. C.; Tang, Y., Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **2009,** *326* (5952), 589-92.
21. Kitagawa, M.; Ara, T.; Arifuzzaman, M.; Ioka-Nakamichi, T.; Inamoto, E.; Toyonaga, H.; Mori, H., Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Res.* **2005,** *12* (5), 291-9.
22. Liscombe, D. K.; Louie, G. V.; Noel, J. P., Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat. Prod. Rep.* **2012,** *29* (10), 1238-50.
23. Huitt-Roehl, C. R.; Hill, E. A.; Adams, M. M.; Vagstad, A. L.; Li, J. W.; Townsend, C. A., Starter unit flexibility for engineered product synthesis by the nonreducing polyketide synthase PksA. *ACS Chem Biol* **2015,** *10* (6), 1443-9.
24. Martin, J. L.; McMillan, F. M., SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **2002,** *12* (6), 783-93.
25. Mui Chan, C.; Zhou, C.; Brunzelle, J. S.; Huang, R. H., Structural and biochemical insights into 2'-O-methylation at the 3'-terminal nucleotide of RNA by Hen1. *Proc Natl Acad Sci U S A* **2009,** *106* (42), 17699-704.
26. Cakici, O.; Sikorski, M.; Stepkowski, T.; Bujacz, G.; Jaskolski, M., Crystal structures of NodS N-methyltransferase from Bradyrhizobium japonicum in ligand-free form and as SAH complex. *J. Mol. Biol.* **2010,** *404* (5), 874-89.
27. Hardwicke, M. A.; Rendina, A. R.; Williams, S. P.; Moore, M. L.; Wang, L.; Krueger, J. A.; Plant, R. N.; Totoritis, R. D.; Zhang, G.; Briand, J.; Burkhart, W. A.; Brown, K. K.; Parrish, C. A., A human fatty acid synthase inhibitor binds beta-ketoacyl reductase in the keto-substrate site. *Nat. Chem. Biol.* **2014,** *10* (9), 774-9.
28. Maier, T.; Leibundgut, M.; Ban, N., The crystal structure of a mammalian fatty acid synthase. *Science* **2008,** *321* (5894), 1315-22.

29. Peng, Y.; Sartini, D.; Pozzi, V.; Wilk, D.; Emanuelli, M.; Yee, V. C., Structural basis of substrate recognition in human nicotinamide N-methyltransferase. *Biochemistry* **2011,** *50* (36), 7800-8.
30. Regueira, T. B.; Kildegaard, K. R.; Hansen, B. G.; Mortensen, U. H.; Hertweck, C.; Nielsen, J., Molecular basis for mycophenolic acid biosynthesis in Penicillium brevicompactum. *Appl. Environ. Microbiol.* **2011,** *77* (9), 3035-43.
31. Yeh, H.; Chang, S. L.; Chiang, Y. M.; Bruno, K. S.; Oakley, B. R.; Wu, T. K.; Wang, C. C., Engineering Fungal Nonreducing Polyketide Synthase by Heterologous Expression and Domain Swapping. *Org. Lett.* **2013,** *15* (4), 756-759.
32. Qian, J.; Khandogin, J.; West, A. H.; Cook, P. F., Evidence for a catalytic dyad in the active site of homocitrate synthase from Saccharomyces cerevisiae. *Biochemistry* **2008,** *47* (26), 6851-8.
33. Khaleeli, N.; Busby, R. W.; Townsend, C. A., Site-directed mutagenesis and biochemical analysis of the endogenous ligands in the ferrous active site of clavaminate synthase. The His-3 variant of the 2-His-1-carboxylate model. *Biochemistry* **2000,** *39* (29), 8666-73.
34. Kyte, J.; Doolittle, R. F., A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982,** *157* (1), 105-132.
35. Conrad, R. S.; Massey, L. K.; Sokatch, J. R., D- and L-isoleucine metabolism and regulation of their pathways in Pseudomonas putida. *J. Bacteriol.* **1974,** *118* (1), 103-11.
36. Steyn, P. S.; Vleggaar, R., Biosynthesis of Asteltoxin by Cultures of Emericella-Variecolor - the Role of Propionate in the Biosynthesis and Evidence for a 1,2-Bond Migration in the Formation of the Bistetrahydrofuran Moiety. *Journal of the Chemical Society-Chemical Communications* **1984,** (15), 977-979.
37. Steyn, P. S.; Vleggaar, R.; Wessels, P. L., Biosynthesis of the Aurovertin-B and Aurovertin-D - the Role of Methionine and Propionate in the Simultaneous Operation of 2 Independent Biosynthetic Pathways. *Journal of the Chemical Society-Perkin Transactions 1* **1981,** (4), 1298-1308.
38. Vagstad, A. L.; Bumpus, S. B.; Belecki, K.; Kelleher, N. L.; Townsend, C. A., Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *J. Am. Chem. Soc.* **2012,** *134* (15), 6865-77.
39. Helfrich, E. J.; Piel, J., Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **2016,** *33* (2), 231-316.
40. Skiba, M. A.; Sikkema, A. P.; Fiers, W. D.; Gerwick, W. H.; Sherman, D. H.; Aldrich, C. C.; Smith, J. L., Domain Organization and Active Site Architecture of a Polyketide Synthase C-methyltransferase. *ACS Chem Biol* **2016**.
41. Poust, S.; Phelan, R. M.; Deng, K.; Katz, L.; Petzold, C. J.; Keasling, J. D., Divergent mechanistic routes for the formation of gem-dimethyl groups in the biosynthesis of complex polyketides. *Angew. Chem. Int. Ed. Engl.* **2015,** *54* (8), 2370-3.
42. Krissinel, E.; Henrick, K., Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2256-68.

43. Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J., Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **1993,** *229* (1), 105-24.
44. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012,** *9* (7), 671-5.
45. Waterman, D. G.; Winter, G.; Parkhurst, J. M.; Fuentes-Montero, L.; Hattne, J.; Brewster, A.; Sauter, N. K.; Evans, G., The DIALS framework for integration software. *CCP4 Newsletter on protein crystallography* **2013,** *49*, 16-19.
46. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 125-32.
47. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 213-21.
48. Karplus, P. A.; Diederichs, K., Linking crystallographic model and data quality. *Science* **2012,** *336* (6084), 1030-3.
49. Sheldrick, G. M., A short history of SHELX. *Acta Crystallogr A* **2008,** *64* (Pt 1), 112-22.
50. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658-674.
51. Cowtan, K., Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 4), 470-8.
52. Cowtan, K., Fitting molecular fragments into electron density. *Acta Crystallogr D Biol Crystallogr* **2008,** *64* (Pt 1), 83-9.
53. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 4), 486-501.
54. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007,** *372* (3), 774-97.
55. Kleywegt, G. J.; Jones, T. A., Model building and refinement practice. *Methods Enzymol.* **1997,** *277*, 208-30.
56. Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.8. 2015.
57. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011,** *7*, 539.
58. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A., Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012,** *28* (12), 1647-9.

Chapter 3: Exploring fungal polyketide *C*-methylation through combinatorial domain swaps

3.1. Introduction

The adaptation of biological catalysts to perform desired reactions on synthetic scales – biocatalysis – has led to greener and more efficient chemical pathways, several of which are vital for pharmaceutical and industrial products.[1] A typical biocatalyst is an enzyme or organism derived directly from a natural source and/or modified to better suit the intended application, maximizing properties such as stability, activity, and regio- or stereoselectivity. Mankind has employed biocatalysts since the dawn of agriculture through the use of native yeast strains to leaven bread and ferment wine, beer, and mead. While these techniques were improved empirically over the centuries, the modern field of biocatalysis began within the last century with the application of biological catalysts to specific chemical transformations, notably the production of stereopure mandelonitrile and hydroxylated steroids.[1, 2] Since then, biocatalysis has been categorized into three "waves" that sequentially capture the technologies used to produce novel catalysts.[1, 3, 4] Briefly, the first wave includes catalysts that are used as found in nature or with minimal optimization, the second wave took advantage of early protein engineering techniques and rational mutagenesis from enzyme structures, and the third wave partnered directed evolution with high throughput analysis, returning to the empiricism found at the roots of biocatalysis.

A survey of the commercial biocatalysts highlights several common themes and suggests where future efforts might be directed. The most successful class of biocatalysts

has be the ketoreductases (KREDs), which have been used to stereospecifically reduce ketones in the production of pharmaceuticals including statins and antiretrovirals.[1] Transaminases, hydrolases, and oxidases have also served as commercially successful biocatalysts. Powerful and profitable as they have been, these enzymes share reactivity towards activated carbon centers or heteroatoms, but enzymes that catalyze C-C bond formation are noticeably rare despite the importance of such bonds and bond-forming reactions in synthetic organic chemistry.

Polyketide synthases (PKSs) have been long-standing targets for engineering biosynthetic pathways because of the inherent structural and stereochemical complexity built into their products.[5] Like fatty acid synthases (FASs) and non-ribosomal peptide synthetases (NRPSs), PKSs tether growing intermediates to the enzyme by thioesterification to the phosphopantetheine arm of a carrier protein that delivers the intermediates to client domains for successive rounds of elongation and modification. FASs and PKSs share the core reaction – ketosynthase-catalyzed C-C bond formation by decarboxylative Claisen condensation, extending the intermediate by two carbons each cycle. However, PKSs occasionally implement a secondary C-C bond forming reaction through regiospecific α-methylation by an *S*-adenosylmethionine (SAM)-dependent *C*-methyltransferase (CMeT), either free-standing or embedded in the PKS.

This second transformation includes several features that may be desireable for enzyme engineering and biocatalysis. First, the methyl acceptor substrates for PKS CMeTs are surprisingly varied. Bacterial Type II PKS natural products can be *C*-methylated by stand-alone CMeTs following aromatization or cyclization, as in the geminal bismethylation seen in resistomycin and benastatin biosynthesis in

Streptomyces.[6, 7] Type I highly-reducing PKSs (HR-PKSs) found in both marine cyanobacteria and fungi, modular and iterative systems respectively, can singly methylate polyene β-ketoacyl substrates as in curacin and lovastatin biosynthesis, through use of a PKS-embedded CMeT.[8, 9] For both examples, multiple β-ketoacyl species of different chain length could be presented to the CMeT, but only a single regiospecific methylation is observed. Some fungal non-reducing PKSs (NR-PKSs) also contain embedded CMeT domains and are known to methylate linear substrates ranging from C_4 to C_{14} prior to cyclization by the product template domain, with many of these CMeTs acting multiple times during iterative catalysis.[10, 11] That such different acceptor substrates are tolerated across all PKS CMeTs may reflect an inherently pliable active site and substrate flexibility, amenable to engineering for biocatalysis.

Though structures of PKS CMeTs remain few in number compared to established families of biocatalysts, two features common to all CMeTs may be advantageous to future efforts to engineer general C-C bond forming biocatalysts. In Type I PKS CMeTs, there is a two subdomain architecture comprised of an *N*-terminal cap and *C*-terminal Rossmann fold, with the active site positioned in a cleft at their interface.[9, 10] The presence of the *N*-terminal subdomain is a diagnostic indicator that the CMeT is active and not an evolutionary structural remnant, as in the Ψ-CMeT in mammalian FAS with only a vestigial helical bundle. The majority of small molecule methyltransferases (MTases) are Class I, or Rossmann fold-containing, and predominantly act on more nucleophilic *O*-, *N*-, or *S*-methyl acceptors and have little or no *N*-terminal subdomain.[12, 13] The CurJ and PksCT CMeT active sites are large and relatively hydrophobic, even though some methylated substrates are as small as acetoacetyl-ACP, though no acceptor

substrates have been observed in CMeT active sites, only the β-ketoacyl moiety appears to make specific contacts with binding pocket residues due to the lack of H-bond donors and acceptors other than the catalytic His-Glu dyad.[10] Therefore, the *N*-terminal subdomain and large acceptor binding pockets of PKS CMeTs could be attributes for modulating selectivity or activity towards non-native substrates.

Finally, recent advances in synthetic SAM analogues has greatly expanded the repertoire of alkyl substituents that can be transferred by MTases and opened new avenues for C-C bond biocatalysis.[14] Instead of a methyl group, these analogs contain *n*-alkyl, allyl, azido, alkynyl, or benzyl groups at the sulfonium center.[15, 16] Significant effort is still required to translate MTases into general alkylation biocatalysts, however.

While a free-standing, "plug-and-play" collection of CMeTs would be ideal, an ensemble of regiospecific CMeTs that could be added to an engineered polyketide biosynthetic pathway would also be useful. Previous work in the Townsend lab has leveraged the ability to deconstruct multienzymes into their constituent domains as a way to explore NR-PKS activity. Chain length, cyclization, and product release have been surveyed by domain deconstruction and combinatorial domain swap reactions in prior work.[17-19] We sought to expand this approach to study fungal polyketide *C*-methylation. A deeper understanding of how fungal NR-PKS CMeTs act in a programmed fashion – to methylate only specific positions on the acceptor substrate during specific iterative cycles – might facilitate rational regiospecific methylation of engineered polyketides. Lessons learned through this endeavor will explore the "rules" of NR-PKS CMeTs and serve as guidelines for subsequent efforts to employ this highly specific family of alkylating enzymes and, hopefully, accelerate the development of C-C bond-forming biocatalysts.

3.2. Results and Discussion

3.2.1. Preparation of Group VI and VII NR-PKS domain fragments

In order to build a domain library for combinatorial swap experiments, we surveyed the literature for fungal Group VI and Group VII NR-PKSs.[11] Only those PKSs that had a known product associated with them were included, to allow for comparison of the native methylation pattern to the pattern seen with *in vitro* domain swap reactions. Several of the PKSs selected were only associated with immediate post-PKS products, while others are also associated with a final natural product. Finally, we omitted PKSs that utilized complex starter units derived from upstream HR-PKSs or FASs such as the asperfuranone PKSs AfoG and AfoE and the alkylisoquinoline synthase PkiA.[11, 20] The candidates selected were PksCT (*Monascus purpureus*)[21], MpaC (*Penicillium brevicompactum*)[22], DtbA (*Aspergillus niger*)[23], and PkeA, AusA, and PkbA (*A. nidulans*).[11] Each of these PKSs utilizes an acetyl starter unit, produces tetra- or pentaketides, and installs one to three *C*-methyl groups in the native product, and we reasoned that their similarity would maximize the chances for successful non-cognate interactions (Figure 3.1).

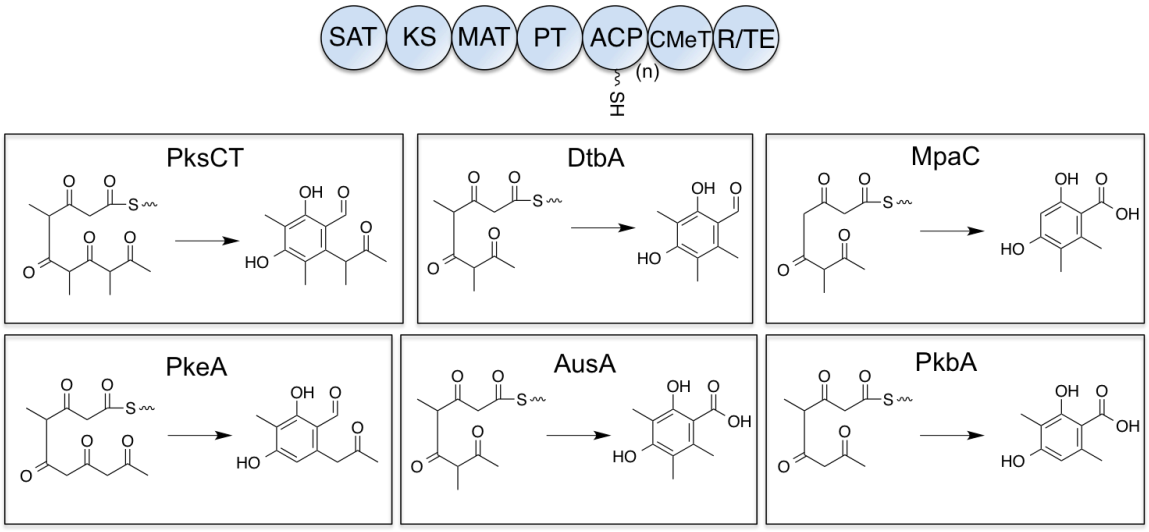


Figure 3.1: Fungal NR-PKSs selected for domain deconstruction and combinatorial swaps. Group VI and VII NR-PKSs share a common domain architecture (top), with Group VI terminating in a TE domain to release the polyketide as the carboxylic acid and Group VII terminating in a R domain to give an aldehyde. Some PKSs, including DtbA, contain two ACP domains. The curated group of NR-PKSs is shown in boxes with the linear methylated intermediate shown on the left of the arrow and the released product on the right. (SAT: starter unit-ACP transacylase, KS: ketosynthase, MAT: malonyl-CoA:ACP transacylase, PT: product template, ACP: acyl carrier protein, CMeT: *C*-methyltransferase, R: reductase, TE: thioesterase).

Using the Udwary-Merski algorithm in combination with the early experience of dissecting PksCT domains (see Chapter 2)[10, 24], we generated expression constructs for the SAT-KS-MAT tridomain, and CMeT and ACP mono-domains (Appendix B). These three would constitute the "minimal methylating PKS," with all necessary domains to carry out multiple rounds of extension and methylation. Based on the initial PksCT results, we decided to omit PT and R or TE domains and relied on spontaneous intramolecular C-O cleavage of the *holo*-ACP thioester to release intermediates as pyrones, a commonly observed PKS derailment product. Test expressions showed that not all desired fragments expressed or were soluble, especially for the SAT-KS-MAT fragments; this outcome is consistent with previous attempts to express SAT-KS-MATs across multiple fungal NR-PKS Groups and did not improve when alternate cut sites were used (Table 3.1).[19]

NR-PKS	SAT-KS-MAT	$ACP_{(n)}$	CMeT
PksCT	Expression, soluble	Expression, soluble	Expression, soluble
DtbA	No expression	Expression, soluble	Expression, soluble
MpaC	No expression	No expression	No expression
AusA	No expression	Expression, soluble	Expression, soluble
PkbA	No expression	No expression	Expression, very unstable
PkeA	Expression, soluble	Expression, soluble	Expression, soluble

Table 3.1: Domain deconstruction of Group VI & VII NR-PKSs.

Unfortunately, SAT-KS-MAT fragments for PkeA and PksCT were the only two available for experiments, despite several attempts to optimize cut-sites and expression conditions for the remaining SAT-KS-MAT fragments. Another notable result of our domain deconstruction was the complete lack of soluble fragments from MpaC. We also found expressing fragments from PkbA challenging, but were able to generate small amounts of the CMeT monodomain. Overall, both ACP and CMeT monodomain fragments were generally well behaved for most of the PKSs in our library. The success rate of domain deconstruction highlights several bottlenecks to this approach, namely the necessary removal of introns from the expression sequence and the difficulty of heterologous expression for proteins of high molecular weight. While identifying incorrectly annotated exon boundaries can be done by mRNA sequencing if producing conditions in the native organism are known, we did not obtain experimental exon boundaries for cryptic PKSs activated by *en masse* promoter replacement strategies.[11] *In silico* approaches like the exon predictor FGENESH can be useful if there is high confidence in a single alternate exon boundary, especially if it restores highly conserved

residues as in the case of PksCT. However, multiple predicted coding sequences can impose a significant obstacle to identifying the true coding sequence.

3.2.2. In vitro *CMeT domain swaps result in altered methylation patterns*

From the soluble, expressed fragments we were able to conduct exploratory reactions to gauge the cross-compatibility of CMeTs with components from other NR-PKSs. All *apo*-ACPs were activated using the promiscuous phosphopantetheinyl-transferase Sfp from *Bacillus subtilis* to give *holo*-ACP.[25] PKS domain fragments were incubated with acetyl-SNAC, malonyl-SNAC, and SAM, and reactions were extracted and analyzed by HPLC and UPLC-ESI-MS. Reconstitution of the minimal PksCT was discussed in detail in Chapter 2, and the minimal PkeA reaction demonstrated a similar outcome (Figure 3.2A). The major product was the unmethylated triketide **1**, however the tetraketide **2** was also detected as a minor product. Thus, the spontaneous release of unmethylated intermediates – either from the *holo*-ACP thioester or the KS active site thioester – appears to be a general trend for CMeT-containing NR-PKSs. We speculate that tetraketide **2** may be favored in PkeA because the native substrate is only methylated once following the third extension (see Figure 3.1), so longer unmethylated substrates must be better tolerated for productive biosynthesis, compared to PksCT. Intramolecular pyrone release still strongly competes with extension in the absence of a methylated substrate.

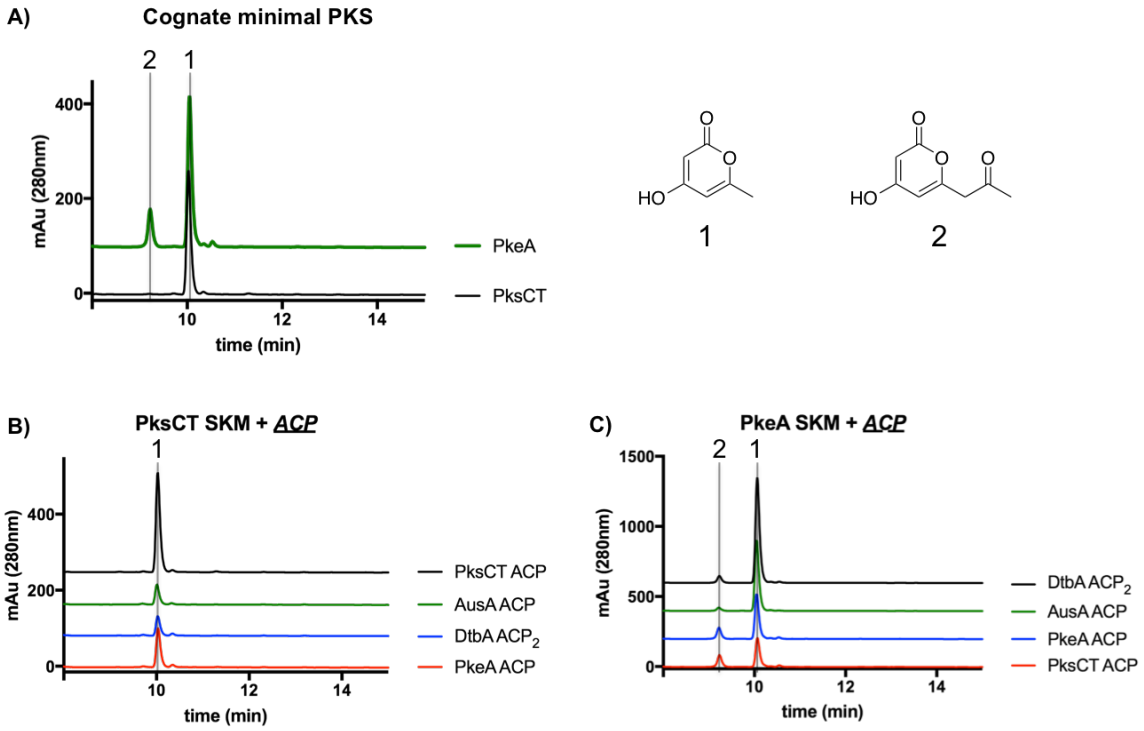


Figure 3.2: Reconstitution of the "minimal PKS" for PksCT and PkeA and ACP swaps. A) HPLC traces respectively, monitored at 280 nm, for PksCT and PkeA minimal PKS reactions. While PksCT only produces the common PKS derailment product triketide **1**, PkeA also produces a minor product tetraketide **2**, perhaps reflecting an inherent tolerance for unmethylated substrates seen in native PkeA biosynthesis. B) Non-cognate ACPs are competent but less active when paired with PksCT SAT-KS-MAT. C) Non-cognate ACPs are competent and somewhat more active when paired with PkeA SAT-KS-MAT, though the ratio of **1** to **2** varies depending on the ACP. Traces are stacked vertically for clarity. See Appendix B for further characterization.

Subsequently, we swapped ACPs to these minimal PKS reactions to determine the compatability of non-cognate ACPs to participate in acyl substrate transfer and elongation (Figure 3.2B and C). As observed for non-methylating NR-PKSs, ACPs are generally interchangeable, though the yield or ratio of pyrone products depends on the identity of the ACP.[19] While ACPs monodomains tend to be amenable to domain dissection, it is difficult to assess the degree to which it impacts synthesis in this case as the corresponding experiment with full-length PKSs remains out of reach. In non-methylating NR-PKSs, dissection reduced yield by an order of magnitude, but the

differences in product profiles seen in the above ACP swaps are consistent with previous work.[19]

We then added CMeT monodomains to PksCT or PkeA minimal PKSs. Unfortunately, PkbA CMeT was particularly unstable, and after multiple attempts, could only be prepared for a single reaction with PksCT. From the observed pyrones, several aspects of *C*-methylation were evident. Non-cognate CMeTs generally are capable of intercepting acyl-*holo*-ACP species and dramatically reduce the yield of unmethylated polyketides. Because these pyrones have very similar UV-Vis absorption spectra, we assume peak areas can be roughly correlated. In PksCT-based reactions (Figure 3.3A), the triketides **3**, **4**, and **5** accumulate when non-cognate CMeTs are added and are significantly less abundant than **1** in the absence of a CMeT. The position of the methyl group for the singly methylated triketides **3** and **5** has been determined by comparison to an authentic standard of **5** (See Chapter 2 for detailed discussion of **3**).

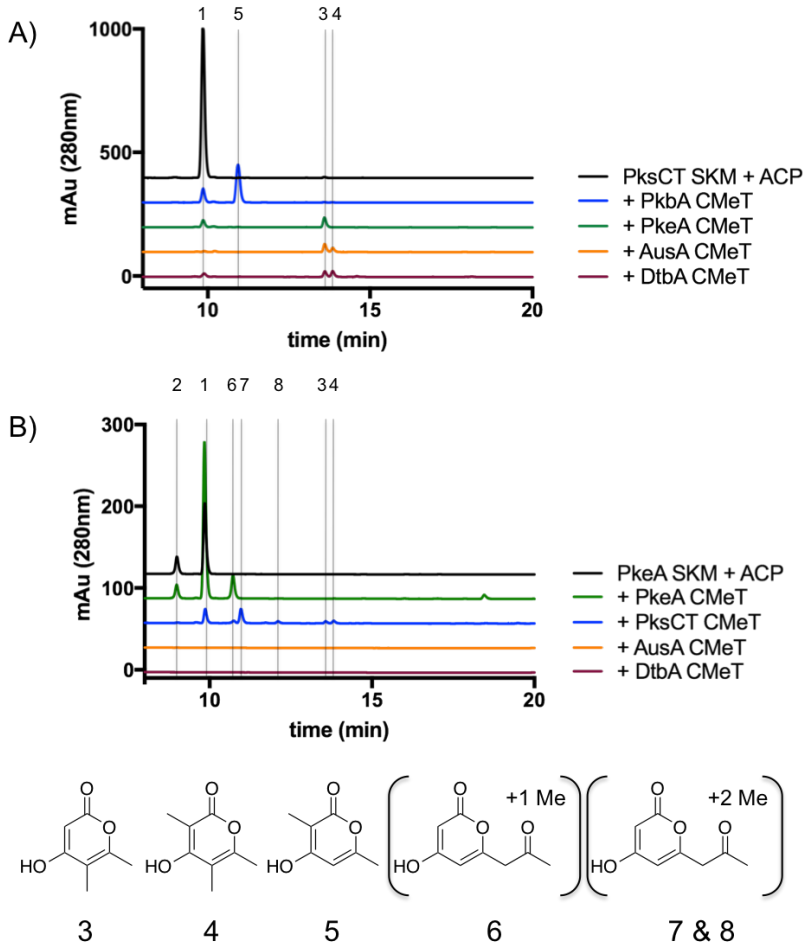


Figure 3.3: CMeT domain swaps result in differently methylated pyrone products. Addition of non-cognate CMeTs to the PksCT (A) or PkeA (B) minimal PKS results in methylated pyrone product profiles that are CMeT-specific. See Appendix B for further characterization.

In PkeA-based reactions (Figure 3.3B), non-cognate CMeTs resulted in no observed product (AusA and DtbA), or an assortment of low-abundance methylated pyrones (PksCT). Notably, the addition of PkeA CMeT to the cognate PkeA SAT-KS-MAT + ACP, increased the amounts of **1** and produced the singly methylated tetraketide **6**, consistent with the native PkeA methylation pattern. However, the methylated positions for tetraketides **6**, **7**, and **8** could not be unambiguously determined from fragmentation during mass spectrometry or comparison to standards. The two doubly methylated tetraketides indicate that PksCT CMeT occassionally skipped methylation

following at least one of the first three extension cycles. The KS domain of PksCT was previously shown to be sensitive to the absence of methylation on the growing intermediate (Chapter 2). Because the native product of PkeA is not methylated following the first two extensions, the KS may be sensitive to the presence of methylation on early stage intermediates. Without additional SAT-KS-MAT fragments to test, the generality of KS sensitivity to methylation remains unclear.

A second set of reactions was also carried out, analogous to those above, but the ACP and CMeT were cognate and paired with a non-cognate SAT-KS-MAT (Figure 3.4).

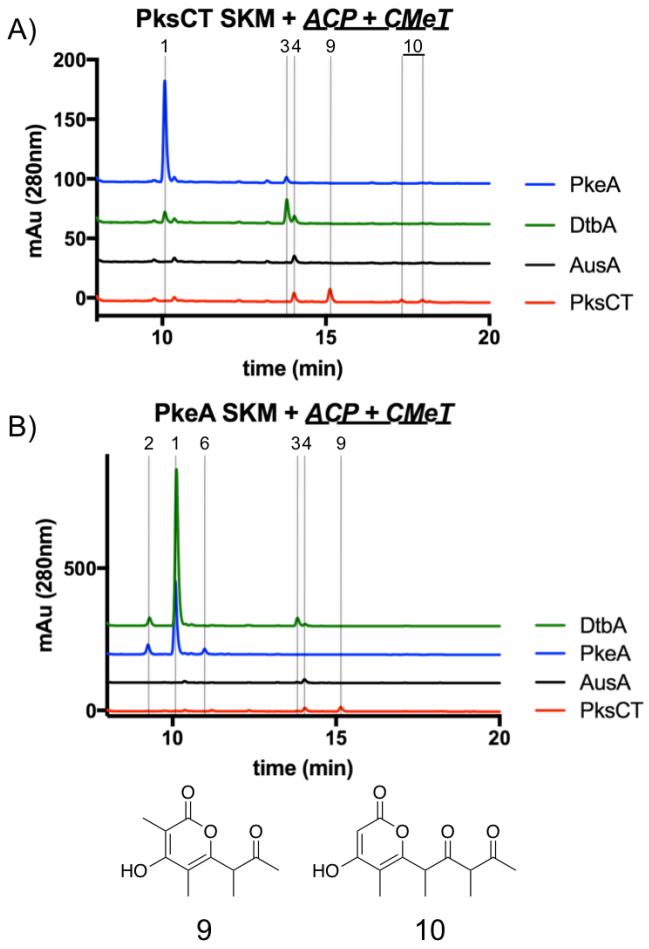


Figure 3.4: Product profiles change slightly when the ACP is cognate to the CMeT. A) PksCT SAT-KS-MAT or B) PkeA SAT-KS-MAT was added to cognate pairs of ACP and CMeT. Extension and methylation depend on binding acyl-ACP intermediates, and ACP identity may influence competiting rates.

In most reactions, the observed products were consistent whether the ACP was paired with the SAT-KS-MAT or CMeT, with some exceptions. Paired PkeA ACP and CMeT were less effective at producing methylated pyrones with PksCT SAT-KS-MAT than when the ACP was paired with the tridomain, perhaps because PkeA CMeT does not natively methylate early acyl intermediates, and spontaneous release at the triketide stage is faster than the methylation programmed by PkeA CMeT. When PkeA SAT-KS-MAT was used, ACP and CMeT pairs from DtbA and AusA were able to generate products, instead of completely stalling extension. PksCT ACP and CMeT were also able to generate the late-stage tetra- and pentaketide pyrones **9** and **10**.

In total, this series of observations reflects the *in trans* nature of acyl-ACP species with both the CMeT and the other client domains, particularly the KS in domain-dissected reconstitution reactions. The growing acyl intermediates must leave the KS active site, travel through solvent while bound to the ACP, and finally dock within the CMeT active site. That the CMeT lowers the rate of extension so significantly suggests that it competes with the KS for binding of acyl-ACP intermediates. In non-methylating NR-PKSs, extension is very rapid and the growing poly β-ketone intermediate likely remains shielded in the KS, as almost no products of intermediate chain length are detected.[26] The addition of a CMeT obviously slows extension as the population of intermediate acyl-ACP species is now split among one additional competing binding partner.

The methylation program for each CMeT appears to be intrinsic, even in the context of domain dissection. The non-cognate CMeT tends to methylate substrates with less fidelity than cognate systems, but with a bias towards the native methylation pattern

of the CMeT. For example, most of the non-cognate CMeTs added to minimal PksCT did not completely eliminate production of the unmethylated triketide **1**, even though the cognate PksCT CMeT completely eliminates this product (see Chapter 2). There remain significant gaps in rationalizing all aspects of the product profile. Without more domain fragments, especially SAT-KS-MATs, it is unclear how sensitive KS domains might be to alternate methylation patterns. The lack of improperly methylated intermediates in PksCT reconstitution suggests reasonably high fidelity to the "programmed" methylation, but not all reaction outcomes are equally informative. For example, if methylating acetoacetyl-ACP after the first Claisen condensation renders the intermediate an unfavorable substrate for continued cycles in PkeA, available free *holo*-ACP will decrease and such an intermediate is less susceptible to spontaneous release (no pyrone formation is possible before the triketide). Attempts to release small intermediates from the ACP by transthioesterification to cysteamine were unsuccessful.[27]

These experiments also lack the cyclization and release domains – PT and R or TE, respectively. Little is known about how these domains interact with their substrates and whether the correct methylation pattern is necessary for activity. Sensitivity of the KS and PT to the stereochemistry of methyl groups (before cyclization), is also unknown. Domain dissection certainly increases the likelihood of racemization of a given α-methyl position as the acyl-ACP docks to other domains *in trans*, a slower, solvent-exposed process. The interdependence of catalytic activity and substrate affinities in interative PKS domains stems from billions of years of co-evolution and will likely remain a fundamentally unresolved problem in the field until large numbers of multidomain structures are available, with multiple bound substrates.

3.2.3. Hypermethylation of linear substrates at high CMeT concentration

The molecular basis of programmed methylation patterns is still incomplete, but some attempts have been made to functionally account for how iterative catalysis operates. Studies on the fungal HR-PKS LovB, responsible for lovastatin biosynthesis, recently probed how acyl-ACP intermediates are appropriately directed towards the next catalytic domain in the programmed sequence.[28] A diene triketide is α-methylated by a CMeT prior to ketoreductase (KR)-catalyzed reduction of the β-ketone to the corresponding alcohol. Because the α-position is most acidic at the β-keto thioester stage, and most available for methylation, the CMeT must act first. As in the NR-PKSs discussed above, LovB appears to be sensitive to substrates that are not methylated, as in the absence of SAM, intermediates are spontaneously released as penta-, hexa-, and heptaketide pyrones.[8] By providing synthetic substrates that varied chain length and β-functional groups, the authors were able to demonstrate that the CMeT has high catalytic efficiency towards only the on-path diene triketide, whereas the KR is less selective and less efficient. The proper order of methylation and reduction is functionally enforced by *C*-methylation being very fast for only a single intermediate while ketoreduction is slower but for a wider range of substrates, consistent with only a single methylation but multiple ketoreductions during synthesis of the LovB product.

We wanted to probe whether a similar process might be followed in methylating NR-PKSs, but by balancing rates of extension and methylation instead of methylation and reduction. A panel of specific synthetic intermediates could not be screened due to their inherent reactivity, so the relative amount of CMeT was varied to influence the overall rate of methylation (Figure 3.5).

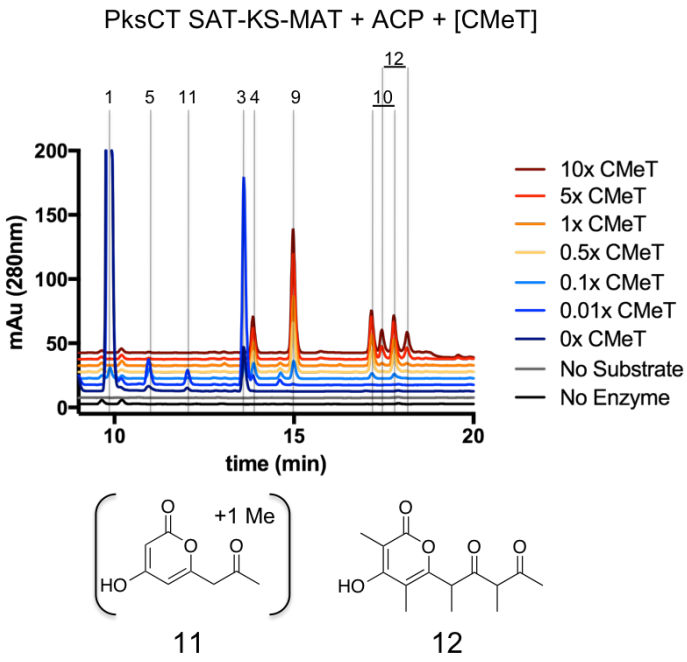


Figure 3.5: Increasing relative amounts of CMeT results in hypermethylation of pentaketide intermediates. PksCT SAT-KS-MAT and ACP (10 μM each) was incubated with increasing concentrations of PksCT CMeT, showing that with excess CMeT, non-native methylation of the final α-position can occur giving the pentaketide **12**. Traces are vertically stacked and the peak for **1** is truncated for clarity. See Appendix B for further characterization.

When even very small amounts of CMeT were added to the reaction, the production of triketide **1** was dramatically reduced, and as the relative concentration of CMeT increased the total number of singly methylated pyrones **3**, **5**, and **11** increasd. At approximately equimolar CMeT, multiply methylated pyrones **4**, **9**, and **10** dominated the reaction profile, consistent with expected on-path intermediates. And with excess CMeT, a new pair of product peaks appeared, proposed to be the hypermethylated pentaketide **12**. Because **12** elutes as two peaks with identical mass, fragmentation, and chromophores, just like on-path pentaketide **10**, it is likely to retain the stereocenters that give rise to diastereomers – ruling out *gem*-dimethylation of a single position and therefore supporting non-native methylation following the final extension.

We propose that the kinetic balance of extension and methylation determines the observed methylation pattern (Figure 3.6). Like the model described for LovB, different domains may have catalytic efficiency towards different substrates such that only the programmed outcome is kinetically favorable. At low concentrations of CMeT, not all programmed methylations can occur before the KS performs another extension generating an off-path intermediate. At high concentration, the selectivity against positions that are not part of the "methylation program" can be overcome by increasing the overall rate of methylation. These experiments cannot replicate the kinetic advantage found in an intact NR-PKS, where the effective concentrations of acyl-ACP and client domains are high. It is also possible that other conformational changes transmitted within an intact NR-PKS could be involved in control of methylation.

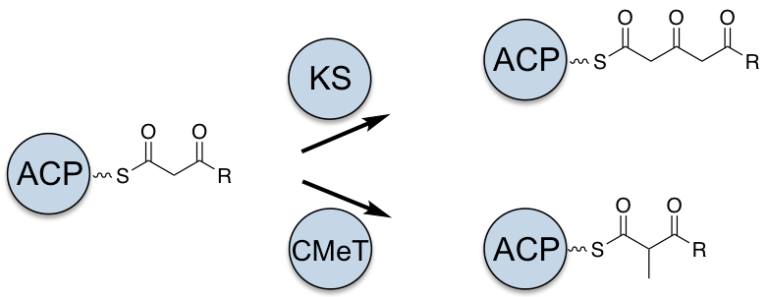


Figure 3.6: Kinetic balance of methylation and extension. Following extension, the acyl-ACP intermediate is a potentially suitable substrate for both methylation and another round of extension. The relative rates of these competing reactions for the various ACP-bound intermediates present during iterative biosynthesis may dictate the "program."

3.3. Conclusions

The development of engineered biosynthetic pathways, and biological catalysts generally, would benefit from regiospecific or programmable C-C bond forming reactions. Fungal PKSs are a source of embedded CMeTs that perform such programmed

alkylations, and little is known about how this process in controlled. We assembled a small library of CMeTs and expanded an *in vitro* domain swap approach to explore the factors controlling polyketide *C*-methylation. We found that NR-PKS CMeTs were capable of methylating acyl-ACP intermediates *in trans*, and that the methylation program is intrinsic to the CMeT. The product profiles could be modulated based on the concentration of CMeT, even allowing for non-native methylation. Together the data suggest that part of methylation control is a kinetic balance between the rates of extension and methylation, though a molecular or structural basis for differential rates remains to be established. A more comprehensive collection of CMeT domains, development of chimeric intact NR-PKSs, and structures of multidomain PKS fragments with bound substrates can each offer additional information about methylation programming.

3.4. Experimental

3.4.1. Materials, strains, and genomic DNA purification

All reagents were purchased from Sigma Aldrich (St. Louis, MO) except AcSNAC, which was prepared synthetically, and MalSNAC, which was prepared enzymatically by MatB and purified.[18, 19] Standard molecular biology procedures were used for isolation of DNA and the creation of expression plasmids. Sequencing was performed at the Johns Hopkins University Synthesis and Sequencing Facility (Baltimore, MD). *M. purpureus* NRRL 1596, *Aspergillus niger* NRRL 328, *Aspergillus nidulans* NRRL 194, and *Penicillium brevicompactum* NRRL 2011 were received from the ARS, USDA (Peoria, IL) and cultured on PDA plates. Collected mycelia were flash

frozen, ground under N_2*(l)* and gDNA was purified using the DNeasy Plant Mini kit (Qiagen, Hilden, Germany). Overlap extension polymerase chain reaction using the primers listed in Appendix B was used to amplify exons and assemble intron-free expression inserts, which were ligated into pET-24a (Novagen) with a C-terminal His_6 tag.

3.4.2. Preparation and expression of PKS domain fragments

Details of expression constructs and domain boundaries can be found in Appendix B. Plasmids were transformed to *E. coli* BL21(DE3) by electroporation and maintained on LB agar plates. Expression cultures were grown in LB media at 37 °C to OD_{600} of 0.6. Cultures were cold-shocked in an ice bath for about 1 h, and induced with 1 mM IPTG (GoldBio, St. Louis, MO) overnight at 18 °C, shaking at 225 rpm. Cell pellets were harvested by centrifugation for 15 min, at 4,000 x g, 4 °C and either flash frozen in $N_2(l)$ and stored at -80 °C until use, or resuspended immediately in 5 mL/g cell pellet in lysis buffer (50 mM potassium phosphate, 300 mM NaCl, 10 % glycerol, pH 7.6). Resuspended cells were lysed by sonication on ice for 10 x 10 s at 40% amplitude (Vibra-Cell Ultrasonic Processor, Sonics & Materials, Inc., Newtown, CT). The lysate was cleared by centrifugation for 25 min at 27,000 x g, 4 °C and batch bound to Co^{2+}-TALON (Clontech, Mountain View, CA) at 4 °C for 1 h. Expressed proteins were purified by gravity column chromatography as follows: the protein-bound resin was washed with lysis buffer (10 CV), followed by lysis buffer + 2 mM imidazole (5 CV) and elution with lysis buffer + 100 mM imidazole (5 CV). Domain fragments were dialyzed into reaction

buffer (100 mM potassium phosphate, 5 % glycerol, pH 7.0), concentrated by diafiltration, and used immediately or flash frozen in $N_2(l)$ and stored at -80 °C.

3.4.3. In vitro domain swap reactions and analysis.

Protein concentration was determined by Bradford assay (BioRad, Hercules, CA) in duplicate using bovine serum albumin as a standard. Prior to *in vitro* reactions, ACPs (50 μM) were activated by Sfp (1 μM) with CoASH (100 μM) and $MgCl_2$ (10 mM) in reaction buffer for 1 h at room temperature as previously reported.[19] Reconstitution reactions contained 10 μM of each included domain with 0.5 mM AcSNAC, 2 mM MalSNAC, 2 mM SAM, and 1 mM TCEP in reaction buffer totaling 250 μL. AcSNAC was prepared synthetically and MalSNAC was purified from MatB reactions.[26] After 4 h at room temperature, the reactions were quenched with 5 μL concentrated HCl and extracted into ethyl acetate 3 x 250 μL. The combined organic extracts were dried to a residue and resuspended in 250 μL of 20 % aqueous acetonitrile. Extracts were analyzed on an Agilent 1200 HPLC with autosampler by injecting 100 μL onto a Prodigy ODS3 column (4.5 x 250 mm, 5μ, Phenomenex, Torrence, CA) with 5-85 % MeCN/H_2O, with 0.1 % formic acid, over 40 min at 1 mL/min and monitored at 280 nm. Mass spectrometric analysis was done using a Waters Acquity Xevo G-2 UPLC-ESI-MS in positive ion mode, with 5 μL injected on a BEHC C18 column and 10-90 % MeCN/H_2O, with 0.1 % formic acid, over 10 min.

3.5. References

1. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., Engineering the third wave of biocatalysis. *Nature* **2012,** *485* (7397), 185-94.
2. Sedlaczek, L., Biotransformations of steroids. *Crit Rev Biotechnol* **1988,** *7* (3), 187-236.
3. Arnold, F. H., Combinatorial and computational challenges for biocatalyst design. *Nature* **2001,** *409* (6817), 253-7.
4. Schoemaker, H. E.; Mink, D.; Wubbolts, M. G., Dispelling the myths--biocatalysis in industrial synthesis. *Science* **2003,** *299* (5613), 1694-7.
5. Staunton, J.; Weissman, K. J., Polyketide biosynthesis: a millennium review. *Natural Product Reports* **2001,** *18* (4), 380-416.
6. Ishida, K.; Fritzsche, K.; Hertweck, C., Geminal Tandem C-Methylation in the Discoid Resistomycin Pathway. *J Am Chem Soc* **2007,** *129*, 12648-12649.
7. Schenk, A.; Xu, Z.; Pfeiffer, C.; Steinbeck, C.; Hertweck, C., Geminal bismethylation prevents polyketide oxidation and dimerization in the benastatin pathway. *Angew Chem Int Ed Engl* **2007,** *46* (37), 7035-8.
8. Ma, S. M.; Li, J. W.; Choi, J. W.; Zhou, H.; Lee, K. K.; Moorthie, V. A.; Xie, X.; Kealey, J. T.; Da Silva, N. A.; Vederas, J. C.; Tang, Y., Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **2009,** *326* (5952), 589-92.
9. Skiba, M. A.; Sikkema, A. P.; Fiers, W. D.; Gerwick, W. H.; Sherman, D. H.; Aldrich, C. C.; Smith, J. L., Domain Organization and Active Site Architecture of a Polyketide Synthase C-methyltransferase. *ACS Chem Biol* **2016**.
10. Storm, P. A.; Herbst, D. A.; Maier, T.; Townsend, C. A., Functional and Structural Analysis of Programmed C-Methylation in the Biosynthesis of the Fungal Polyketide Citrinin. *Cell Chem Biol* **2017,** *24* (3), 316-325.
11. Ahuja, M.; Chiang, Y. M.; Chang, S. L.; Praseuth, M. B.; Entwistle, R.; Sanchez, J. F.; Lo, H. C.; Yeh, H. H.; Oakley, B. R.; Wang, C. C., Illuminating the diversity of aromatic polyketide synthases in Aspergillus nidulans. *J Am Chem Soc* **2012,** *134* (19), 8212-21.
12. Martin, J. L.; McMillan, F. M., SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struct Biol* **2002,** *12* (6), 783-93.
13. Liscombe, D. K.; Louie, G. V.; Noel, J. P., Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat Prod Rep* **2012,** *29* (10), 1238-50.
14. Bennett, M. R.; Shepherd, S. A.; Cronin, V. A.; Micklefield, J., Recent advances in methyltransferase biocatalysis. *Curr Opin Chem Biol* **2017,** *37*, 97-106.
15. Dalhoff, C.; Lukinavicius, G.; Klimasauskas, S.; Weinhold, E., Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nat Chem Biol* **2006,** *2* (1), 31-2.
16. Stecher, H.; Tengg, M.; Ueberbacher, B. J.; Remler, P.; Schwab, H.; Griengl, H.; Gruber-Khadjawi, M., Biocatalytic Friedel-Crafts alkylation using non-natural cofactors. *Angew Chem Int Ed Engl* **2009,** *48* (50), 9546-8.

17. Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A., Deconstruction of iterative multidomain polyketide synthase function. *Science* **2008,** *320* (5873), 243-6.
18. Vagstad, A. L.; Newman, A. G.; Storm, P. A.; Belecki, K.; Crawford, J. M.; Townsend, C. A., Combinatorial Domain Swaps Provide Insights into the Rules of Fungal Polyketide Synthase Programming and the Rational Synthesis of Non-Native Aromatic Products. *Angew Chem Int Ed Engl* **2013,** *52*, 1718-1721.
19. Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A., Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming. *J Am Chem Soc* **2014,** *136* (20), 7348-62.
20. Chiang, Y. M.; Szewczyk, E.; Davidson, A. D.; Keller, N.; Oakley, B. R.; Wang, C. C., A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in Aspergillus nidulans. *J Am Chem Soc* **2009,** *131* (8), 2965-70.
21. Shimizu, T.; Kinoshita, H.; Ishihara, S.; Sakai, K.; Nagai, S.; Nihira, T., Polyketide synthase gene responsible for citrinin biosynthesis in Monascus purpureus. *Appl Environ Microbiol* **2005,** *71* (7), 3453-7.
22. Regueira, T. B.; Kildegaard, K. R.; Hansen, B. G.; Mortensen, U. H.; Hertweck, C.; Nielsen, J., Molecular basis for mycophenolic acid biosynthesis in Penicillium brevicompactum. *Appl Environ Microbiol* **2011,** *77* (9), 3035-43.
23. Yeh, H.; Chang, S. L.; Chiang, Y. M.; Bruno, K. S.; Oakley, B. R.; Wu, T. K.; Wang, C. C., Engineering Fungal Nonreducing Polyketide Synthase by Heterologous Expression and Domain Swapping. *Org Lett* **2013,** *15* (4), 756-759.
24. Udwary, D. W.; Merski, M.; Townsend, C. A., A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *J Mol Biol* **2002,** *323* (3), 585-98.
25. Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T., Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **1998,** *37* (6), 1585-95.
26. Vagstad, A. L.; Bumpus, S. B.; Belecki, K.; Kelleher, N. L.; Townsend, C. A., Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *J Am Chem Soc* **2012,** *134* (15), 6865-77.
27. Belecki, K.; Townsend, C. A., Biochemical determination of enzyme-bound metabolites: preferential accumulation of a programmed octaketide on the enediyne polyketide synthase CalE8. *J Am Chem Soc* **2013,** *135* (38), 14339-48.
28. Cacho, R. A.; Thuss, J.; Xu, W.; Sanichar, R.; Gao, Z.; Nguyen, A.; Vederas, J. C.; Tang, Y., Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase. *J Am Chem Soc* **2015,** *137* (50), 15688-91.

Chapter 4: *In trans* hydrolysis of carrier protein-bound acyl intermediates by CitA during citrinin biosynthesis

4.1. Introduction

Polyketide synthases (PKSs) are multi-domain biosynthetic enzymes that catalyze the formation of complex carbon scaffolds through repeated rounds of extension and modification of simple acyl substrates.[1-3] Intermediates are covalently tethered to the PKS by thioester linkage to a post-translationally installed phosphopantetheine arm of the acyl carrier protein (ACP) and delivered to the active sites of other domains for programmed chemistry. Mature intermediates are released from the ACP through a number of mechanisms including hydrolysis, cyclization, and reduction, or transfered to a downstream carrier protein. The reactive nature of many PKS intermediates, however, occasionally results in the presence of acyl-*holo*-ACP species that are not on-path to the programmed PKS product and must be removed for productive chemistry to resume. This reactivity is fundamentally problematic for non-reduced polyketide intermediates that juxtapose nucleophilic and electrophilic moieties susceptible to intramolecular reactions that derail the acyl intermediate from the programmed path, for example, in the formation of the off-path SEK4 and SEK4b derailment products.[4, 5]

Ensuring that the genetic and metabolic investment in such large biosynthetic machinery is not perpetually waylaid, several different strategies to remove unproductive acyl intermediates have been employed across multiple types of PKSs. Many Type I PKSs have a thioesterase (TE) domain at their *C*-terminus commonly associated with product release, and some also have hydrolytic activity towards other acyl-*holo*-ACP

species.[6] In some cases, however, this hydrolytic activity has been maintained by other proteins acting *in trans*. Fungal non-reducing PKSs (NR-PKSs) lacking a *C*-terminal TE domain can have a separate metallo-β-lactamase type TE in the same gene cluster that hydrolyzes the mature ACP-bound intermediate as in atrochrysone carboxylic acid biosynthesis.[7]

In bacterial trans-AT PKSs, PedC was recently shown to serve as a stand-alone acyl hydrolase (AH) capable of liberating several acyl species from the ACP in pederin biosynthesis.[8] The presence of hydrolysis mechanisms across disparate PKSs suggests that this activity may represent a general strategy to maintain PKS efficiency in the absence of a *cis*-TE.

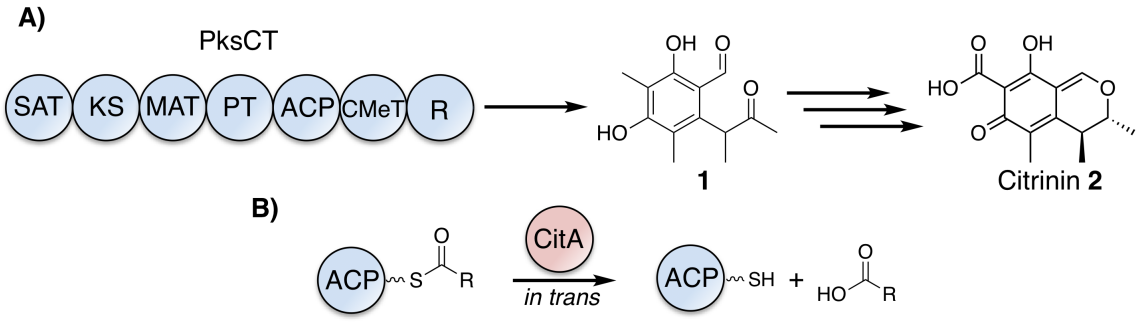


Figure 4.1: Proposed role of CitA in citrinin biosynthesis. A) Domain architecture of PksCT, the fungal NR-PKS responsible for the first isolable intermediate in citrinin biosynthesis. The benzaldehyde **1** is subsequently reduced, cyclized, and oxidized by downstream enzymes CitB, CitC, CitD, and CitE to afford citrinin **2**. B) CitA is proposed to intercept ACP-bound acyl intermediates *in trans* and catalyze hydrolysis to release free *holo*-ACP during stalled biosynthesis. (SAT: starter unit-ACP transacylase, KS: ketosynthase, MAT: malonyl-CoA:ACP transacylase, PT: product template, ACP: acyl carrier protein, CMeT: C-methyltransferase, R: reductase).

While many fungal NR-PKSs contain a *C*-terminal TE that may have an editing role, the recently categorized Group VII PKSs terminate instead with a reductase (R) domain that catalyzes NADPH-dependent release of the mature thioester intermediate as an aldehyde (Figure 4.1).[9] Whether R domains are capable of carrying out an analogous editing function is unclear, but such a function would presumably involve the expenditure

of NADPH and be energetically expensive for the producing organism relative to hydrolysis. Additional work has identified the presence of a putative α/β-hydrolase-encoding gene adjacent to Group VII PKSs in a number of biosynthetic gene clusters, and co-expression of the PKS and hydrolase gives greater titers of the post-PKS product compared to expression of the PKS alone.[10, 11] Several roles for such putative hydrolases have been proposed, and this uncertainty was exacerbated by the initial misannotation of the start codon for these genes.[12] While their exact function is not resolved, *in trans* editing of acyl intermediates is a possible role for these accessory proteins and one consistent with the *in vivo* results.

4.2. Results

4.2.1. In vitro reconstitution of PksCT with CitA

We sought to identify the activity of CitA (GenBank: BAE95339), the putative hydrolase adjacent to PksCT, the Group VII NR-PKS in the *Monascus purpureus* gene cluster responsible for citrinin biosynthesis.[13, 14] Previously, CitA was re-annotated from an oxidoreductase to a hydrolase due to an incorrect initial assignment of the start codon, and deletion of CitA in the citrinin-producer *Monascus ruber* greatly decreased, but did not eliminate, production of the post-PKS aldehyde **1**.[10] Additionally, co-expression of both PksCT and CitA in the heterologous host *Aspergillus oryzae* gave significantly higher titers of **1** than the PKS alone. Based only on these observations, however, the role of CitA could not be established definitively. Using a previously reported domain-deconstructed system, we added CitA to reconstituted PksCT *in vitro* to assess its effect on the characterized product profiles from specific domain combinations.[15] Addition of CitA to all combinations of PksCT domains resulted in a concentration-dependent

decrease in the derailment products triketides **3** and **4**, tetraketide **5**, and pentaketide **6**, as well as the on-path product aldehyde **1** (Figure 4.2). No new products were found as might be expected if CitA catalyzed a modification of PKS-bound intermediates or the final product **1**. Unlike reported results of *in vivo* coexpression of PksCT with CitA, we did not observe any increase in **1** at any concentration of CitA. We suspect that synthetic inefficiencies from domain dissection may explain the lack of increase in **1** *in vitro*. The *in vivo* use of intact PksCT allows for intramolecular interaction between the ACP and other PksCT domains; however, we could not obtain soluble intact PksCT from heterologous expression in *Escherichia coli* or *Saccharomyces cerevisiae* to test this hypothesis.

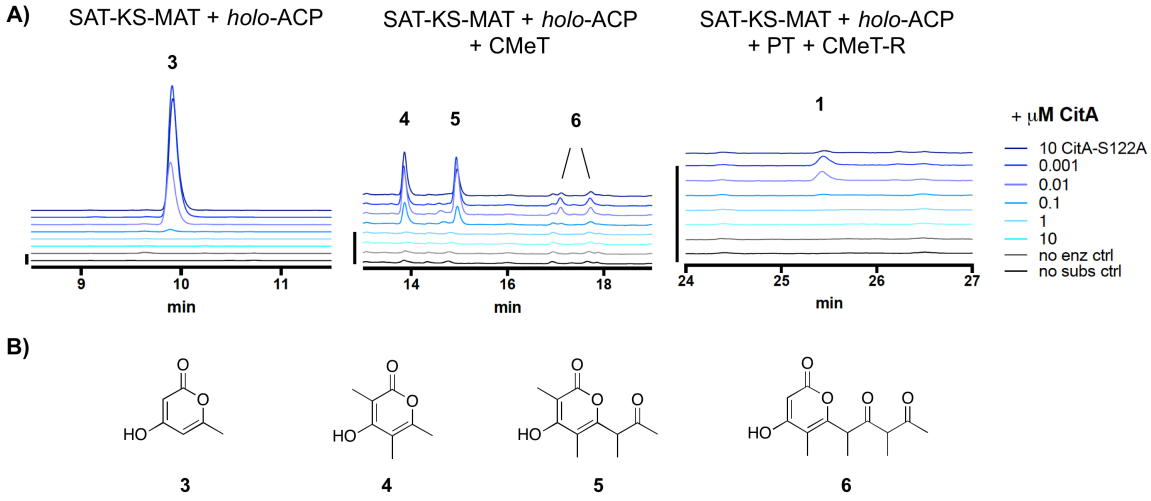


Figure 4.2: *In vitro* reconstitution of PksCT with CitA reduces product yield. A) Increasing concentration of CitA resulted in a corresponding decrease in the amount of all products for all tested domain combinations. Absorbance traces are vertically offset and shown at 280 nm; the bar to the left of each set of traces represents 10 mAu. For a detailed discussion of these derailment products, see Chapter 2. B) Structures of PksCT derailment products detected in partial PksCT reconstitution reactions.

4.2.2. CitA hydrolyzes acyl species from PksCT holo-ACP

Since CitA only appeared to decrease product yields, we speculated that it might be hydrolyzing one or more intermediates, analogous to a recent observation of *in trans*

acyl-*holo*-ACP hydrolysis in pederin biosynthesis.[8] We identified a GxSxG active site motif common to α/β-hydrolases that was conserved in several putative CitA-like hydrolases adjacent to Group VII NR-PKSs, including AfoE, PkdA, PkeA, and PkhA (Figure 4.3).[9, 16, 17] We also generated a homology model of CitA using the CPHmodels 3.2 server, which identified the yeast serine hydrolase FSH1/YHR049W as the closest homolog of known structure (PDB: 1YCD).[18, 19] The homology model suggested a Ser122-His235-Asp207 catalytic triad, consistent with recent *in vivo* observations of the CitA homolog MppD in azaphilone biosynthesis in *M. purpureus*.[11]

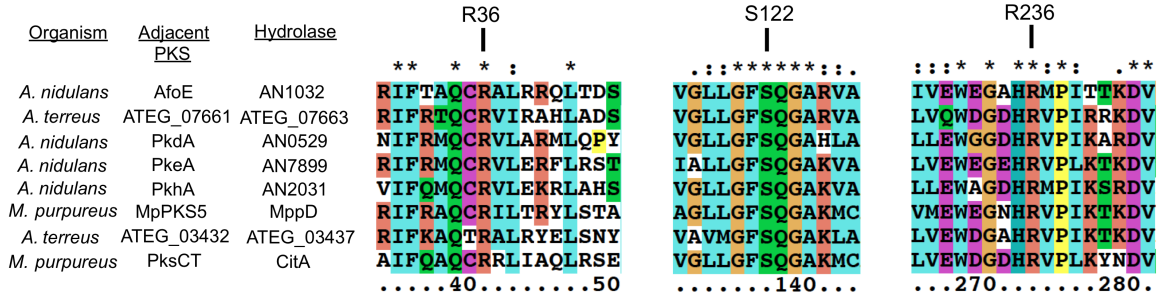


Figure 4.3: Alignment of CitA to putative hydrolases adjacent to Group VII NR-PKSs. Eight Group VII NR-PKSs with known products were found to have a putative hydrolase gene adjacent to the PKS. Alignment of the hydrolase sequences identified a strongly conserved active site motif around Ser122. We also identified two conserved basic residues – Arg36 and Arg 236 – located near the active site entrance, based on homology modeling (discussed below). The alignment was done using ClustalX v2.1.[20]

To test CitA hydrolysis against the simplest of acyl-ACP species, PksCT *apo*-ACP (ACP_{CT}) monodomain was activated by the promiscuous phosphopantetheinyl-transferase Sfp with [1-^{14}C]-acetyl-CoA to give the radiolabeled acetyl-*holo*-ACP_{CT} and incubated with CitA or CitA-S122A. Following separation of the reaction products by SDS-PAGE, we observed that the acetyl radiolabel was lost from ACP_{CT} in a CitA-dependent fashion, but CitA-S122A was inert (Figure 4.4). A radiolabeled band consistent with the larger CitA was not detected in the gel, suggesting that CitA does not

retain the acetyl species for transfer to a downstream acceptor but rapidly hydrolyzes it to free acetate.

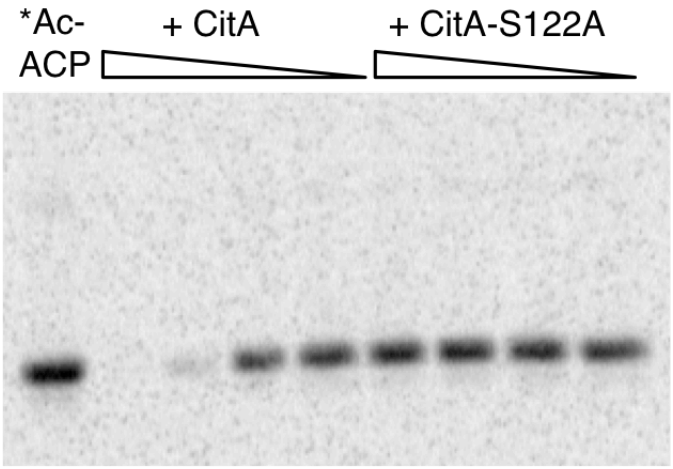


Figure 4.4: CitA hydrolyzes acetyl-*holo*-ACP to free *holo*-ACP and acetate. Radiolabeled [1-^{14}C]-acetyl-*holo*-ACP was incubated with increasing concentrations of CitA or the putative active site mutant CitA-S122A. CitA removed the radiolabel in a concentration-dependent fashion, but CitA-S122A was inert. No corresponding transfer of the radiolabel to CitA was detected, suggesting it is not involved in acyl transfer.

Previous examples of editing TEs or hydrolases in PKS biosynthesis have been shown to have varying degrees of substrate promiscuity towards both on-path and likely off-path intermediates.[6, 8] Many of the acyl-ACP_{CT} intermediates en route to **1** are not accessible due to their inherent reactivity, preventing analysis of late-stage tri-, tetra-, or pentaketide intermediates. However, we tested the hydrolytic activity of CitA against malonyl-*holo*-ACP_{CT} and acetoacetyl-*holo*-ACP_{CT} – acyl species bound to ACP_{CT} at early stages of the biosynthetic cycle. The mass of the acyl-ACPs was determined by UPLC-ESI-MS before and after incubation with CitA (Figure 4.5 and Appendix C). *apo*-ACP_{CT} was observed as a quartet of masses consistent with oxidation, acetylation, and both modifications that was also observed in every ACP_{CT} species throughout these experiments.[21] Loss of mass consistent with acyl hydrolysis was observed for both malonyl and acetoacetyl substrates when incubated with CitA.

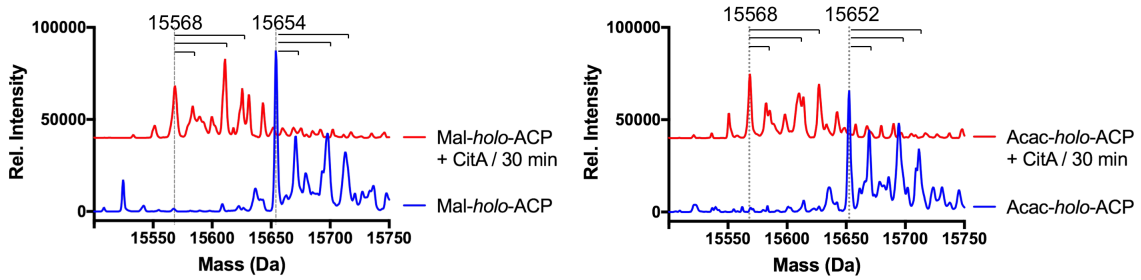


Figure 4.5: Cit A hydrolyzes malonyl and acetoacetyl intermediates from *holo*-ACP$_{CT}$. ACP$_{CT}$ was activated by Sfp to give Mal-*holo*-ACP$_{CT}$ (left, 15654 Da) or Acac-*holo*-ACP$_{CT}$ (right, 15652 Da) and then incubated with CitA for 30 min. UPLC-ESI-MS analysis showed masses consistent with the loss of each acyl species to give free *holo*-ACP$_{CT}$ (15568 Da). The ACP purifies as a mixture of four species: unmodified, oxidized [+16 Da], acetylated [+42 Da], and both [+58 Da], and these species are indicated by the horizontal bars extending from the unmodified mass (dashed vertical lines).

Competition experiments combining malonyl- and acetoacetyl-*holo*-ACP$_{CT}$ did not show a significant substrate preference by CitA (See Appendix C). Attempts to obtain Michaelis-Menten values to compare these acyl species using the *N*-acetylcysteamine (SNAC) or CoA thioesters as substrate mimics were unsuccessful due to an inability to saturate CitA activity (Figure 4.6). The weak activity against small molecule substrates suggested that the ACP is an essential component of CitA substrates, focusing CitA activity towards PKS-bound intermediates and limiting indiscriminant hydrolysis of acyl-CoA species.

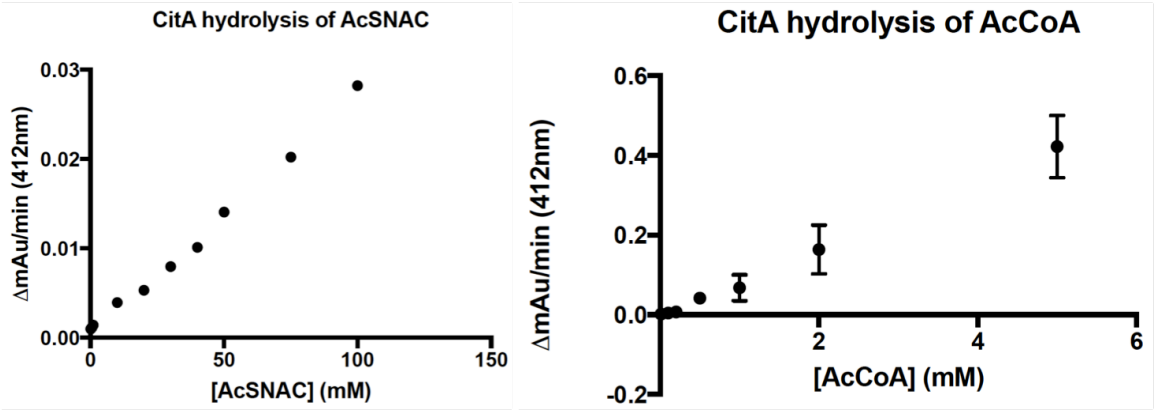


Figure 4.6: Example Michaelis-Menten kinetics of hydrolysis by CitA. Using Ellman's reagent to detect and quantify release of free thiol following AcSNAC or AcCoA hydrolysis, CitA activity could not be saturated, even at high substrate concentrations.

4.2.3. Identifying key in trans *interactions between PksCT ACP and CitA*

Several investigations into the interactions of ACP domains with PKS or fatty acid synthase client domains, including product template, ketosynthase, and acyltransferase domains, have identified conserved basic residues that form contacts with the phosphodiester moiety of the phosphopantetheine arm or conserved acidic residues on the acyl-*holo*-ACP.[22-26] Using our homology model in conjunction with sequence alignments of CitA to related hydrolases, we identified two basic residues – Arg36 and Arg236 – near the entrance of the CitA active site that could be involved in the CitA:ACP_{CT} binding interface (Figure 4.7). Both are predicted to be within ~20 Å of Ser122, approximately the length of the extended phosphopantetheine arm, placing them in close proximity to negatively charged residues on the ACP surface.

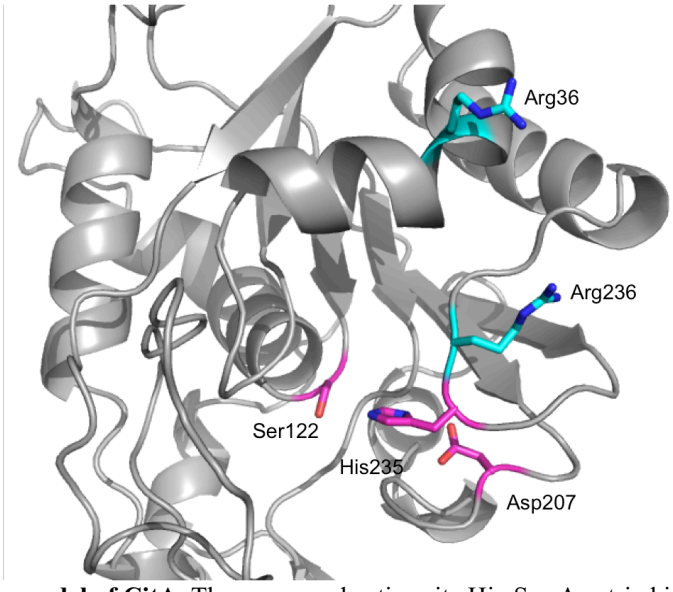


Figure 4.7: Homology model of CitA. The conserved active site His-Ser-Asp triad is shown in magenta. Two conserved basic residues, Arg36 and Arg236, are within ~20 Å of Ser122 which is the approximate length of the phosphopantetheine arm. For details on building the homology model, see 4.5 Experimental and Appendix C.

CitA-R36A and CitA-R236A mutants were generated and incubated with [1-^{14}C]-acetyl-*holo*-ACP_{CT} in a time-course assay and separated by SDS-PAGE (Figure 4.8). CitA-R236A was indistinguishable from the wild type and both completely hydrolyzed the acetyl radiolabel. In contrast, CitA-R36A had minimal hydrolytic activity suggesting that Arg36 has an essential role in mediating the *in trans* interaction between acyl-*holo*-ACP_{CT} and CitA. The conserved nature of this basic residue among CitA homologs in other fungal NR-PKS biosynthetic gene clusters reinforces the conclusion that CitA-like hydrolases preferentially act upon PKS-bound acyl intermediates.

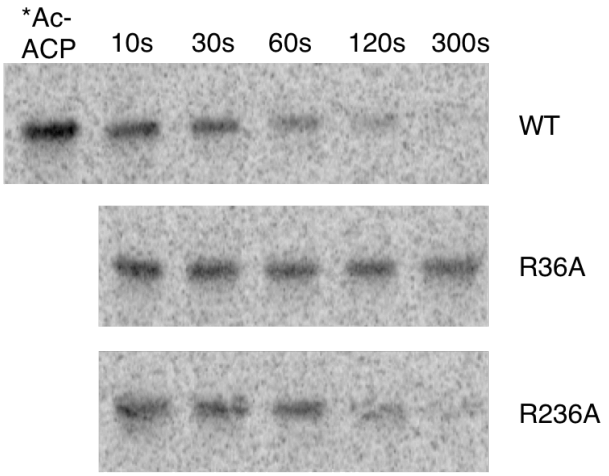


Figure 4.8: The CitA:ACP interface. Time-course incubation of CitA (1 μM) with [1-^{14}C]-acetyl-*holo*-ACP_{CT} (10 μM) shows significantly reduced hydrolysis by the R36A mutant compared to wild type. The R236A mutation had minimal effect on CitA hydrolysis.

4.3. Discussion

Unlike the clean linearity shown in textbook descriptions of biosynthetic pathways, production of secondary metabolites is subject to multiple competing processes. Deciphering contributions of both on- and off-path reactions can be particularly difficult for PKSs (as well as fatty acid synthases and non-ribosomal peptide synthetases) because intermediates are covalently tethered to the phosphopantetheine

arm. Premature hydrolysis of the labile thioester is a general off-path process, as is improper ordering of β-tailoring steps like reduction or dehydration with the next round of extension. A final common competing process in PKSs is spontaneous decarboxylation of an enzyme-bound malonyl unit to give a non-productive acetyl unit. Because multi-domain structural information for PKSs with bound intermediates is only just becoming available, very little is known about the changes that can occur due to these off-path processes and how they are detected and resolved.

The observed *in trans* hydrolytic activity of CitA exemplifies two aspects of polyketide biosynthesis that converge to restore PKS activity and have become more widely appreciated in recent years: *in trans* domain:domain interactions and editing of non-productive intermediates. Early studies on modular PKSs supported an assembly-line model of biosynthesis where each module extended and modified the tethered substrate.[27] As genomic sequencing efforts accumulated thousands of uncharacterized biosynthetic gene clusters, increasingly diverse variations of this model were identified. A significant number of bacterial PKSs adopt a semi-modular architecture, where extension and tailoring steps are encoded within the domains of the module, but the extender units (malonyl or methylmalonyl) are loaded onto ACPs by a free-standing acyltransferase domain or *trans*-AT.[28] The same *trans*-AT typically serves each module in the PKS. Fungal PKSs have also been shown to have *trans*-acting domains, notably LovC, a *trans* enoylreductase (ER) in lovastatin biosynthesis which substitutes for an inactive *cis*-ER in the highly-reducing PKS.[29] Finally, thioesterase function in Group V NR-PKSs, which lack a *C*-terminal TE, has been maintained *in trans* by β-metallolactamase-like proteins. It remains unclear why *trans*-acting domains are used in some biosynthetic pathways but

not others. However, a genetic duplication resulting in a separate *trans* domain could allow for mutations to alter the function of the new monodomain.

Editing of PKS-bound intermediates is also a deviation from classical co-linearity seen in early PKS research. Like with *trans*-acting domains, examples of editing by hydrolysis of acyl intermediates are found in both bacterial and fungal PKSs. The TE domain from PksA in *Aspergillus parasiticus* was shown to hydrolyze acetyl and hexanoyl, but not malonyl, substrates.[6] Since then, a surprising breadth of activities has been observed in fungal NR-PKS TEs, including deacetylation[30] and enol lactonization[31], but the generality of editing activity among *cis*-TEs is not known. The closest comparison to CitA, however, was recently found in bacterial *trans*-AT PKSs during pederin biosynthesis.[8] Identified as an acyl hydrolase (AH), the investigators found that PedC, a protein encoded in the pederin cluster homologous to the *trans*-AT PedD, was active against a number of acyl-ACP substrates, particularly short chain acetyl and butanoyl thioesters.

Therefore, we propose that CitA serves a role at the confluence of these biosynthetic strategies. We suspect that the intramolecular relationship between acyl-ACP intermediates and client domains favors on-path biosynthesis, but when derailment occurs (e.g. non-productive decarboxylation of malonyl-ACP) and the PKS stalls, CitA can intercept and clear the ACP-bound acyl species (Figure 4.9). The relative promiscuity of CitA towards ACP-bound acyl substrates can be rationalized in two ways. First, several off-path events could occur that result in stalled PksCT, including the aforementioned decarboxylation of malonyl-ACP, loss of KS-bound acyl intermediates with malonyl-ACP unable to complete extension, or insufficient levels of SAM to

methylate the acetoacetyl-ACP intermediate. Longer chain length intermediates may also be substrates for CitA but could not be assayed experimentally because spontaneous intramolecular C-O cleavage to the corresponding pyrone prohibits loading them onto an ACP, or even synthesizing them. Secondly, by strongly preferring ACP-bound intermediates, futile hydrolysis of small molecule acyl thioesters (e.g. AcCoA) is disfavored. It is still unknown how specific CitA-like editing hydrolases are towards their native PKS, and follow-up experiments pairing Group VII NR-PKSs with non-cognate hydrolases (including distant relatives such as PedC).

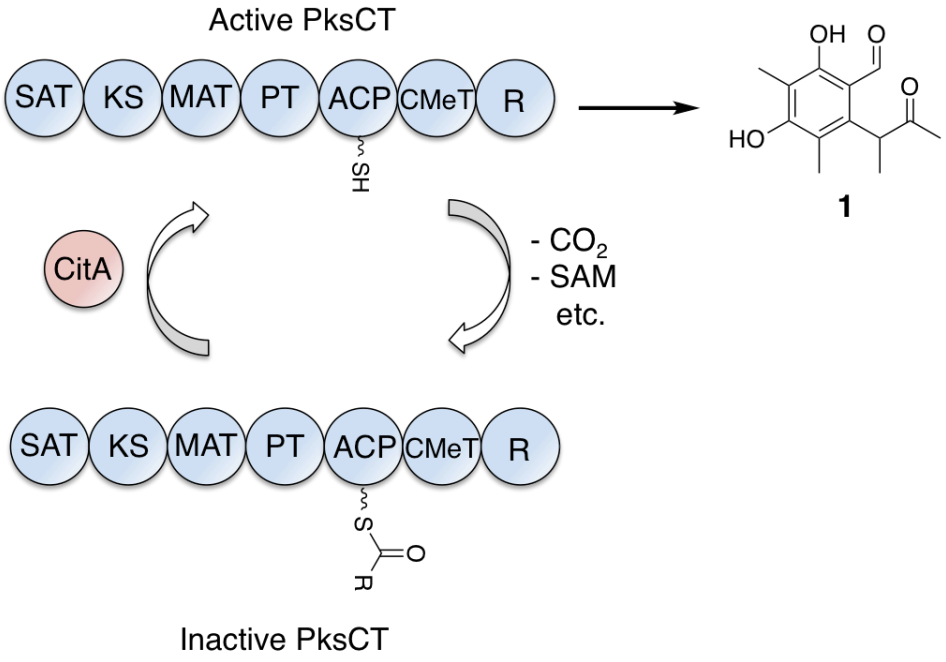


Figure 4.9: CitA intercepts stalled PksCT and restores production by clearing acyl-ACP species. Through a number of spontaneous processes, the PKS can become stalled by having an ACP-bound acyl group that cannot participate in on-path biosynthesis. By removing the blocking acyl group, CitA frees PksCT to resume production of the aldehyde **1**.

Additionally, the convergence of *in vivo* and *in vitro* results remains a priority. While an editing role is consistent with the *in vivo* co-expression of PksCT and CitA giving an approximately 20-fold increase in the product aldehyde **1**, it was surprising to

find no increase *in vitro*.[10] Some possible explanations are discussed above, but a more thorough answer will require use of an intact, full-length PksCT. Unfortunately, expression of full-length fungal PKSs for *in vitro* analysis remains beyond reach. We envision *in trans* editing hydrolases as a potential tool for improving engineered biosynthetic pathways, and any utility would benefit from a deeper understanding of the interactions between these hydrolases and the PKSs they interact with.

4.4. Conclusions

In conclusion, our results expand the scope of *in trans* editing of acyl-ACP intermediates in fungi, and highlight it as a generalized strategy to ensure efficient polyketide biosynthesis, especially when the PKS lacks other self-editing mechanisms. We show that several carrier protein-bound acyl intermediates are removed by CitA and identify conserved features that are essential to CitA:ACP binding and hydrolysis. While acetyl-*holo*-ACP_{CT} is an on-path intermediate to initiate citrinin biosynthesis, hydrolysis of acetyl-*holo*-ACP_{CT} by CitA prevents stalling of PksCT in the event of spontaneous, non-productive decarboxylation of malonyl-*holo*-ACP_{CT} during extension of longer chain-length intermediates. Additionally, hydrolytic activity towards malonyl or acetoacetyl intermediates may be necessary when other client domains like the ketosynthase, *C*-methyltransferase, or reductase have lost their bound co-substrates or have been otherwise inactivated. It remains unclear how PksCT and CitA collaborate in producing organisms, as the *in vitro* data do not replicate the increased yield seen *in vivo*. Acyl-*holo*-ACP_{CT} species may be less exposed in intact PksCT than in our dissected system due to faster intramolecular shuttling of ACP-bound intermediates, favoring binding of ACP_{CT} to other PksCT client domains relative to intermolecular association

with CitA. When PksCT is dissected, all ACP_{CT} interactions occur *in trans*, with hydrolysis by CitA expectedly competing against polyketide synthesis. Further study of relative expression, catalytic rates, precursor availability, and other factors affecting efficiency may uncover the optimal balance between editing and productive biosynthesis. As a significant number of fungal and bacterial PKSs have been found to adopt *in trans* editing mechanisms, future work to engineer these polyketide pathways, or any others lacking metabolically optimized editing, may benefit by including an *in trans* hydrolase, perhaps under distinct and tunable control.

4.5. Experimental

4.5.1. Materials and Cloning

All reagents were purchased from Sigma-Aldrich unless described otherwise. Commercially obtained malonyl-CoA and acetoacetyl-CoA were re-purified to minimize free CoA and acetyl-CoA. Purification was performed on an Agilent 1100 HPLC using a Phenomenex Kinetex C_{18} column (5 μm, 250 x 10 mm) with 100% H_2O + 0.1% TFA for 2 min followed by 0-35% MeCN/H_2O + 0.1% TFA over 23 min at 4 mL/min. Purified CoAs were lyophilized and stored at -80 °C. [1-^{14}C]-acetyl-CoA was purchased from Perkin Elmer and used without further purification. *M. purpureus* NRRL 1596 was obtained from the Agricultural Research Service, USDA. Mycelia were cultured on PDA plates and gDNA was purified using the DNeasy Plant mini Kit (Qiagen). The coding sequence for CitA was amplified from purified gDNA using primers MpOrf4-g5 and MpOrf4-g3 and Phusion polymerase (New England Biolabs) and ligated into pCR-Blunt vector using the Zero Blunt PCR Cloning Kit (Invitrogen). An expression insert

reflecting the re-annotated start codon was amplified from this plasmid using primers MpOrf4-5.2 and MpOrf4-3 and Phusion polymerase. The insert and pET-24a (Novagen) were digested with NdeI and NotI-HF (New England Biolabs) and ligated with T4 ligase (New England Biolabs). CitA mutant expression plasmids were prepared analogously to the wild-type plasmid by overlap extension PCR using the corresponding primers in Appendix C. Expression plasmid sequences were verified at the Johns Hopkins University Synthesis and Sequencing Facility.

4.5.2. In vitro reconstitution of CitA with PksCT

A detailed description of PksCT reconstitution reactions and reaction products was previously published.[15] Briefly, PksCT mono- and multidomain expression plasmids and CitA or CitA mutant expression plasmids were transformed into *E. coli* BL21(DE3). Expression cultures were grown at 37 °C in LB medium to OD_{600} of ~0.6, cooled in an ice bath, and induced with IPTG (1 mM, GoldBio) at 18 °C. Following overnight expression, cells were harvested by centrifugation at 4000 x g and flash frozen in liquid nitrogen for storage at -80 °C or resuspended in 5 mL/g lysis buffer (50 mM K/PO_4, 300 mM NaCl, 10% glycerol, pH 7.5) for immediate purification. The resuspended slurries were sonicated for 10 x 10 s, 40% amplitude (Vibra-Cell Ultrasonic Processor), on ice and cleared by centrifugation at 25,000 x g, 4 °C for 25 min. Lysates were batch bound to Co^{2+}-TALON (Clontech) for 1 h, 4 °C and purified by gravity column, washing with 10 column volumes of lysis buffer and 5 column volumes of lysis buffer + 2 mM imidazole, and eluted with 3 column volumes of lysis buffer + 100 mM imidazole. Purified proteins were dialyzed against reaction buffer (100 mM K/PO_4, 5% glycerol, pH 7.0) flash frozen in liquid nitrogen, and stored at -80 °C until use. Protein concentrations were determined

using the Bradford Assay (Bio-Rad) in duplicate, with bovine serum albumin as a standard. ACP_{CT} was activated by Sfp with CoASH and $MgCl_2$. Reconstitution reactions included the listed PksCT domains at 10 μM each, with increasing concentration of CitA or CitA-S122A, along with 0.5 mM AcSNAC, 2 mM MalSNAC, 2 mM SAM, 1 mM NADPH, and 1 mM TCEP in reaction buffer to 250 μL. Reactions were run for 4 h at 25 °C before being acidified with 5 μL concentrated HCl, extracted into ethyl acetate 3 x 250 μL, and dried. The residue was dissolved in 250 μL 20% ACN/H_2O and analyzed on an Agilent 1200 HPLC using a Phenomenex Prodigy ODS3 column (5 μm, 4.5 x 250 mm) with 5-85% MeCN/H_2O + 0.1% formic acid for 40 min at 1 mL/min. Products were detected by diode array at 280 nm.

4.5.3. Radiolabel and LCMS assays of CitA activity

For radiochemical assays, purified *apo*-ACP_{CT} (50 μM) was activated with [1-^{14}C]-acetyl-CoA (100 μM) by Sfp (1 μM) in dialysis buffer supplemented with $MgCl_2$ (10 mM) for 1.5 h at room temperature. As shown in Figure 4.4, radiolabeled acetyl-*holo*-ACP_{CT} (10 μM) was incubated with CitA or CitA-S122A (10, 1, 0.1, or 0.01 μM) in reaction buffer (50 μL total) for 5 min at 25 °C. As shown in Figure 4.8, radiolabeled acetyl-*holo*-ACPCT (10 μM) was incubated with CitA, CitA-R36A, or CitA-R236A (1 μM), with aliquots taken at the given time-points. Reactions were quenched with 5X SDS loading buffer (lacking β-mercaptoethanol) and separated by SDS-PAGE (16%). Gels were dried, exposed to a phosphorimager screen overnight, and scanned on a Typhoon 9410 Variable Mode Imager (Amersham Biosciences). Data were processed and analyzed using ImageJ.[32] For LCMS assays, *apo*-ACPCT (100 μM) was activated with malonyl-CoA or acetoacetyl-CoA (200 μM) by Sfp (1 μM) in reaction buffer supplemented with

$MgCl_2$ (10 mM) for 2 h at room temperature. Acyl-*holo*-ACP_{CT} (50 μM) was incubated with CitA (1 μM) for 30 min at 25 °C, and 2.5 μL were diluted to 200 μL with 20 % MeCN/H_2O. The reactions were analyzed on a Waters ACQUITY Xevo G-2 UPLC-ESI-MS in positive ion mode by injecting 5 μL onto a BEH C_4 column with 100% H_2O + 0.1% formic acid for 1 min followed by 0-80% MeCN/H_2O + 0.1% formic acid over 6.5 min. ACP_{CT} eluted at 6 min and detected ions at 600-2200 m/z were analyzed in MassLynx using MaxEnt deconvolution to give intact protein masses.

4.5.4. Kinetics experiments with CitA and small molecule substrate mimics

Crystalline AcSNAC or HPLC purified AcCoA were dissolved in dialysis buffer to the desired stock concentration. CitA (1 μM) was preincubated with Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid), 0.2 mM) at 25 °C for 2 min. Absorbance at 412 nm was monitored using a Cary UV/Vis spectrophotometer upon addition of small molecule substrate at 3 s intervals for 3 min at 25 °C. Absorbance data were analyzed in Prism 7 and fit using a linear regression for the first 15 s of collected data.

4.5.5. CitA homology model

The amino acid sequence for CitA was submitted to the CPHmodel server v3.2 and a model was built from PDB: 1YCD with a Z-score of 30.4. The homology model was submitted to the RAMPAGE server and contained five residues in outlier regions, three on the boundary of allowed conformations and two residues on loops distant from the active site. See Appendix C for detailed homology model results.

4.6. References

1. Staunton, J.; Weissman, K. J., Polyketide biosynthesis: a millennium review. *Natural Product Reports* **2001,** *18* (4), 380-416.
2. Crawford, J. M.; Townsend, C. A., New insights into the formation of fungal aromatic polyketides. *Nat. Rev. Microbiol.* **2010,** *8* (12), 879-89.
3. Hertweck, C., The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.* **2009,** *48* (26), 4688-716.
4. Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A., Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming. *J. Am. Chem. Soc.* **2014,** *136* (20), 7348-62.
5. Fu, H.; Hopwood, D. A.; Khosla, C., Engineered biosynthesis of novel polyketides: evidence for temporal, but not regiospecific, control of cyclization of an aromatic polyketide precursor. *Chem. Biol.* **1994,** *1* (4), 205-10.
6. Vagstad, A. L.; Bumpus, S. B.; Belecki, K.; Kelleher, N. L.; Townsend, C. A., Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *J. Am. Chem. Soc.* **2012,** *134* (15), 6865-77.
7. Awakawa, T.; Yokota, K.; Funa, N.; Doi, F.; Mori, N.; Watanabe, H.; Horinouchi, S., Physically discrete beta-lactamase-type thioesterase catalyzes product release in atrochrysone synthesis by iterative type I polyketide synthase. *Chem. Biol.* **2009,** *16* (6), 613-23.
8. Jenner, M.; Afonso, J. P.; Kohlhaas, C.; Karbaum, P.; Frank, S.; Piel, J.; Oldham, N. J., Acyl hydrolases from trans-AT polyketide synthases target acetyl units on acyl carrier proteins. *Chem Commun (Camb)* **2016,** *52* (30), 5262-5.
9. Ahuja, M.; Chiang, Y. M.; Chang, S. L.; Praseuth, M. B.; Entwistle, R.; Sanchez, J. F.; Lo, H. C.; Yeh, H. H.; Oakley, B. R.; Wang, C. C., Illuminating the diversity of aromatic polyketide synthases in Aspergillus nidulans. *J. Am. Chem. Soc.* **2012,** *134* (19), 8212-21.
10. He, Y.; Cox, R. J., The molecular steps of citrinin biosynthesis in fungi. *Chem. Sci.* **2016,** *7* (3), 2119-2127.
11. Balakrishnan, B.; Chandran, R.; Park, S. H.; Kwon, H. J., A New Protein Factor in the Product Formation of Non-Reducing Fungal Polyketide Synthase with a C-Terminus Reductive Domain. *J. Microbiol. Biotechnol.* **2015,** *25* (10), 1648-52.
12. Davison, J.; al Fahad, A.; Cai, M.; Song, Z.; Yehia, S. Y.; Lazarus, C. M.; Bailey, A. M.; Simpson, T. J.; Cox, R. J., Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. *Proc Natl Acad Sci U S A* **2012,** *109* (20), 7642-7.
13. Shimizu, T.; Kinoshita, H.; Ishihara, S.; Sakai, K.; Nagai, S.; Nihira, T., Polyketide synthase gene responsible for citrinin biosynthesis in Monascus purpureus. *Appl. Environ. Microbiol.* **2005,** *71* (7), 3453-7.
14. Shimizu, T.; Kinoshita, H.; Nihira, T., Identification and in vivo functional analysis by gene disruption of ctnA, an activator gene involved in citrinin biosynthesis in Monascus purpureus. *Appl. Environ. Microbiol.* **2007,** *73* (16), 5097-103.

15. Storm, P. A.; Herbst, D. A.; Maier, T.; Townsend, C. A., Functional and Structural Analysis of Programmed C-Methylation in the Biosynthesis of the Fungal Polyketide Citrinin. *Cell Chem Biol* **2017,** *24* (3), 316-325.
16. Chiang, Y. M.; Szewczyk, E.; Davidson, A. D.; Keller, N.; Oakley, B. R.; Wang, C. C., A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in Aspergillus nidulans. *J. Am. Chem. Soc.* **2009,** *131* (8), 2965-70.
17. Brenner, S., The molecular evolution of genes and proteins: a tale of two serines. *Nature* **1988,** *334* (6182), 528-30.
18. Nielsen, M.; Lundegaard, C.; Lund, O.; Petersen, T. N., CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.* **2010,** *38* (Web Server issue), W576-81.
19. Quevillon-Cheruel, S.; Leulliot, N.; Graille, M.; Hervouet, N.; Coste, F.; Benedetti, H.; Zelwer, C.; Janin, J.; Van Tilbeurgh, H., Crystal structure of yeast YHR049W/FSH1, a member of the serine hydrolase family. *Protein Sci.* **2005,** *14* (5), 1350-6.
20. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,** *23* (21), 2947-8.
21. Zhang, J.; Sprung, R.; Pei, J.; Tan, X.; Kim, S.; Zhu, H.; Liu, C. F.; Grishin, N. V.; Zhao, Y., Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli. *Mol. Cell. Proteomics* **2009,** *8* (2), 215-25.
22. Weissman, K. J.; Muller, R., Protein-protein interactions in multienzyme megasynthetases. *ChemBioChem* **2008,** *9* (6), 826-48.
23. Barajas, J. F.; Shakya, G.; Moreno, G.; Rivera, H., Jr.; Jackson, D. R.; Topper, C. L.; Vagstad, A. L.; La Clair, J. J.; Townsend, C. A.; Burkart, M. D.; Tsai, S. C., Polyketide mimetics yield structural and mechanistic insights into product template domain function in nonreducing polyketide synthases. *Proc Natl Acad Sci U S A* **2017,** *114* (21), E4142-E4148.
24. Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T., Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proc Natl Acad Sci U S A* **2016,** *113* (7), 1802-7.
25. Nguyen, C.; Haushalter, R. W.; Lee, D. J.; Markwick, P. R.; Bruegger, J.; Caldara-Festin, G.; Finzel, K.; Jackson, D. R.; Ishikawa, F.; O'Dowd, B.; McCammon, J. A.; Opella, S. J.; Tsai, S. C.; Burkart, M. D., Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **2014,** *505* (7483), 427-31.
26. Bruegger, J.; Haushalter, R. W.; Vagstad, A. L.; Shakya, G.; Mih, N.; Townsend, C. A.; Burkart, M. D.; Tsai, S. C., Probing the selectivity and protein.protein interactions of a nonreducing fungal polyketide synthase using mechanism-based crosslinkers. *Chem. Biol.* **2013,** *20* (9), 1135-46.
27. Weissman, K. J.; Leadlay, P. F., Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **2005,** *3* (12), 925-36.
28. Helfrich, E. J.; Piel, J., Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **2016,** *33* (2), 231-316.

29. Ma, S. M.; Li, J. W.; Choi, J. W.; Zhou, H.; Lee, K. K.; Moorthie, V. A.; Xie, X.; Kealey, J. T.; Da Silva, N. A.; Vederas, J. C.; Tang, Y., Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **2009,** *326* (5952), 589-92.
30. Vagstad, A. L.; Hill, E. A.; Labonte, J. W.; Townsend, C. A., Characterization of a fungal thioesterase having Claisen cyclase and deacetylase activities in melanin biosynthesis. *Chem. Biol.* **2012,** *19* (12), 1525-34.
31. Newman, A. G.; Vagstad, A. L.; Belecki, K.; Scheerer, J. R.; Townsend, C. A., Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem Commun (Camb)* **2012,** *48* (96), 11772-4.
32. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012,** *9* (7), 671-5.

Appendix A: Supplementary material for Chapter 2

Table A1: Primers used in this work.

Primer Name	Sequence
MpPksCT-g5	ATGATTGACTCAACTTCGCACTCAAATCTGAG
MpPksCT-g3	TTAATCTAGAAATCCCATGGTCTTCCATGC
MpPksCT-ex1-5	GCATATTAATGATTGACTCAACTTCGCACTC
MpPksCT-ex1-3.2	GTCAACGATGACAGACACATATGCATCAGGGAATGTCTCGAGAACG
MpPksCT-ex2-5.2	CTCGAGACATTCCCTGATGCATATGTGTCTGTCATCGTTGAC
MpPksCT-ex2-3	GCATGCGGCCGCATCTAGAAATCCCATGGTCTTCCATG
MpPksCT-SAT-3	GCATGCGGCCGCTCCACCGAGAAGCTGCCC
MpPksCT-SKM3-3	GCATGCGGCCGCAGAGCGCTGGGATGCGTCTTG
MpPksCT-PT-5	GCATATTAATGACAGCAATTATCGAGGCACC
MpPksCT-PT-3	GCATGCGGCCGCGCCGCTTACCGCAGTAGACG
MpPksCT-ACP-5	GCATATTAATGACACCTCCGACAAAGAAAGCGC
MpPksCT-ACP-3	GCATGCGGCCGCACCGTTACCATTAACGACACCATTAGTGC
MpPksCT-CMeT-5	GCATATTAATGGTCTTGTTTCCTGAGCTAGGGGG
MpPksCT-CMeT-3	GCATGCGGCCGCGTCTGTCAATGGGACCTGCACC
MpPksCT-CMeTAseI-5	CGAGGAGGCCAGACTGATTGATATCAACG
MpPksCT-CMeTAseI-3	CGTTGATATCAATCAGTCTGGCCTCCTCG
MpAct-5	GGAATTCTGCAGATTCTACAACGAACTCCG
MpAct-3	GGAATTCTGCAGTCAGGGAGTTCATAGGAC
MpPksCT-CMeT-Y1955F-5	GTCACTGATGTCTTTGCCAAATCTCC
MpPksCT-CMeT-Y1955F-3	GGAGATTTGGCAAAGACATCAGTGAC
MpPksCT-CMeT-Y1955A-5	GTCACTGATGTCGCGGCCAAATCTCC
MpPksCT-CMeT-Y1955A-3	GGAGATTTGGCCGCGACATCAGTGAC
MpPksCT-CMeT-H2067A-5	CAAATTGTGTGGCGGCGACCCGG
MpPksCT-CMeT-H2067A-3	CCGGGTCGCCGCCACACAATTTG
MpPksCT-CMeT-H2067Q-5	CAAATTGTGTGCAGGCGACCCGG
MpPksCT-CMeT-H2067Q-3	CCGGGTCGCCTGCACACAATTTG

UMA-guided domain dissection

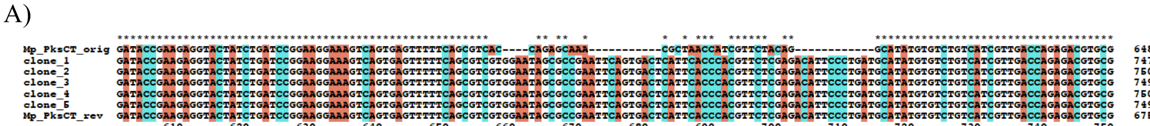
UMAv2.18 was used with the BLOUSUM62 matrix. Secondary structure predictions were generated with ProteinPredict ((Yachdav et al., 2014), https://www.predictprotein.org/). Comparative sequences were curated from NCBI using the Conserved Domain Architecture Retrieval Tool (Geer et al., 2002) with PksCT as the query and included the following entries: XP_001217248, ANID_07903, ANID_00523, CAN87161, XP_001273475, XP_002381902, EHA28237, XP_001818926, XP_001393501, CAK40124, XP_660990.1, XP_658638.1, AAR90253.1, BAE66025.1, XP_001243185.1, CAK48487.1, XP_001559289.1, XP_002149737.1, XP_002567553.1, XP_002487778.1, XP_002340070.1, XP_002384396.1, XP_002848394.1, XP_003070229.1, XP_003010590.1, XP_001395291.2, EHA25844.1, EHA28237.1, and GAA92425.1.

Figure A1: Exon revision and starter unit verification for PksCT.

A) Five independent clones amplified from a cDNA library generated from total RNA under citrinin producing conditions. Each clone exactly matched the FGENESH prediction for the exon 1:exon 2 boundary.

B) Radiolabel assay of acetyl starter unit transfer in PksCT. Lane 1: PksCT SAT, Lane 2: PksCT SAT + ACP without Sfp activation, Lane 3: PksCT SAT + ACP with Sfp activation. NR-PKS ACP domains are often partially phosphopantetheinylated during expression, accounting for the low ACP signal in lane 2.

A)



B)

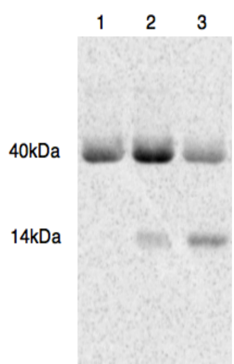


Table A2: Plasmids used in this work:

PksCT domain(s)	Sequence	Plasmid name	Parent vector	Tag	MW (kDa)
SAT	M1-G380	pEMpPksCT-SAT	pET-24a	C-His6	41.8
SAT-KS-MAT	M1-I1288	pEMpPksCT-SKM3	pET-24a	C-His6	144.0
PT	T1285-G1651	pEMpPksCT-PT	pET-24a	C-His6	41.5
ACP	T1652-G1779	pEMpPksCT-ACP	pET-24a	C-His6	15.0
CMeT	V1780-D2182	pEMpPksCT-CMeT	pET-24a	C-His6	46.6
CMeT-Y1955A	V1780-D2182	pEMpPksCT-CMeT-Y1955A	pET-24a	C-His6	46.6
CMeT-Y1955F	V1780-D2182	pEMpPksCT-CMeT-Y1955F	pET-24a	C-His6	46.6
CMeT-H2067A	V1780-D2182	pEMpPksCT-CMeT-H2067A	pET-24a	C-His6	46.6
CMeT-H2067Q	V1780-D2182	pEMpPksCT-CMeT-H2067Q	pET-24a	C-His6	46.6
CMeT-R	V1780-D2593	pEMpPksCT-CMeT-R	pET-24a	C-His6	92.7

Table A3: Crystallographic data collection and refinement statistics.

	CMeT native		**CMeT SeMet**
Data collection			
Space group	H3	H3†	H3†
Cell dimensions			
a, b, c (Å)	98.37, 98.37, 133.43	98.52, 98.52, 133.20	98.13, 98.13, 124.45
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Wavelength (Å)	0.99998	1.90747	0.97929
Swipes	1	2	2
Resolution (Å)	71.80 – 1.65‡	71.85 - 2.05	50.21 – 1.85
R_{merge} (%)*	8.0 (159.6)	8.4 (102.9)	13.5 (274.2)
$I/\sigma I$*	10.61 (1.37)	15.62 (1.38)	13.75 (1.12)
$CC_{1/2}$ (%)*	99.9 (76.0)	99.8 (71.1)	99.9 (57.1)
Completeness (%)*	98.9 (94.3)	97.9 (81.7)	99.8 (99.7)
Redundancy*	6.6 (6.6)	8.5 (2.6)	10.4 (10.2)
Unique reflections*	57,453 (4,045)	59,253 (3,685)	76,184 (5,646)
Refinement			
Resolution (Å)	71.80 – 1.65		
R_{work}/ R_{free}	0.19 / 0.22		
Ramachandran favored / outlier (%)	99.48 / 0.00		
No. atoms	6,512		
Protein	6,173		
Ligand	45		
Solvent	294		
B-factors	36.22		
Protein (Å^2)	36.29		
Ligand (Å^2)	26.84		
Solvent (Å^2)	36.11		
R.m.s deviations			
Bond lengths (Å)	0.009		
Bond angles (º)	0.975		

The resolution cutoff was determined by $CC_{1/2}$ criterion (Karplus and Diederichs, 2012).

*, Highest resolution shell is shown in parenthesis. Datasets of native CMeT were acquired from the same crystal. †Anomalous data processing was applied. ‡A deprecated

conservative resolution cutoff at $I/\sigma I$=2.91 and Rmerge= 69.8% is reached at a resolution of 1.84 Å.

Figure A2: UV-Vis spectra for compounds 2-8.

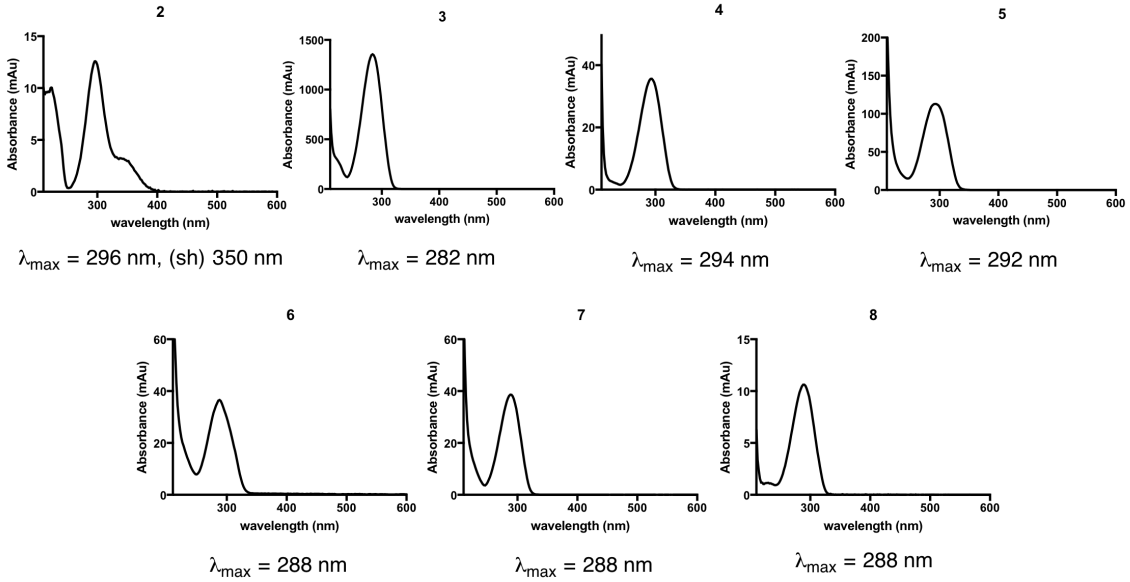
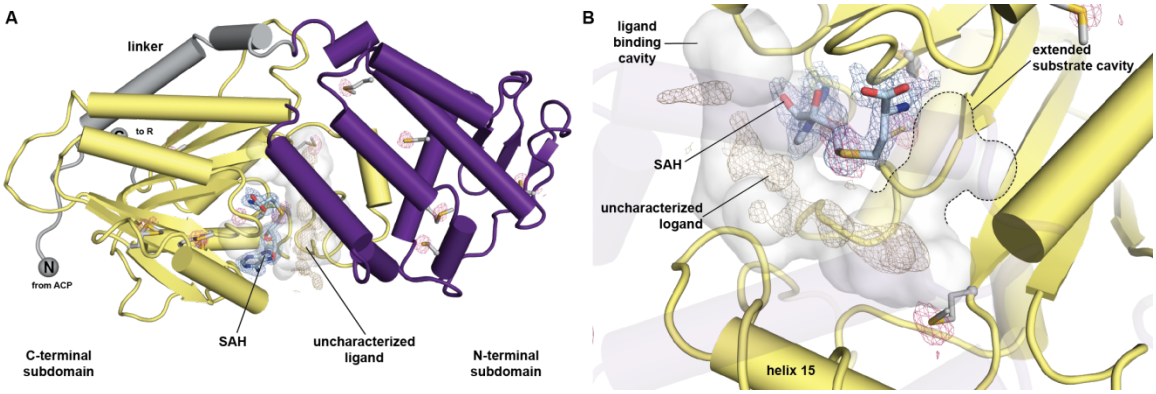


Table A4: Detected masses for compounds 2-8.

#	Calculated [MH+] or [MNa+] m/z	Detected m/z (ppm)	Formula
2	237.1127	237.1121(2.53)	$C_{13}H_{17}O_4$
3	127.0395	127.0391 (3.15)	$C_6H_7O_3$
4	155.0708	155.0704 (2.58)	$C_8H_{11}O_3$
5	211.0970	211.0970 (0.00)	$C_{11}H_{15}O_4$
	233.0787	233.0790 (1.29)	$C_{11}H_{14}O_4Na$
	169.0865	169.0859 (2.37)	$C_9H_{13}O_3[-C_2H_2O]$
6	253.1075	253.1076 (0.40)	$C_{13}H_{17}O_5$
	275.0895	275.0893 (0.73)	$C_{13}H_{16}O_5Na$
	211.0970	211.0965 (2.73)	$C_{11}H_{15}O_4[-C_2H_2O]$
	155.0708	155.0701 (4.51)	$C_8H_{11}O_3[-C_5H_6O_2]$
7	141.0552	141.0553 (0.71)	$C_7H_9O_3$
8	141.0552	141.0552 (0)	$C_7H_9O_3$

Figure A3:. Uncharacterized ligand and anomalous difference density.



(A) The active site tunnel reveals significant F_O-F_C omit difference density at 2.5 σ (brown) with a featured shape, which does not match components of the crystallization condition. Anomalous difference density at 3.7 σ (pink), indicates the presence of sulfur in methionine and cysteine residues as well as in SAH, but no significant difference density as part of the ligand.

(B) The close up view on the active site reveals the expansion of the ligand's difference density along the central tunnel with its end in the junction to the extended substrate cavity. The same contour level of the SAH difference density (blue) suggests either a partial occupancy or various conformations of the less defined ligand. The extent of difference density corresponds approximately to an atomistic binding interface of 500 $Å^2$.

Figure A4: Revised M. purpureus PksCT sequence.

Residues encoded by alternate exon boundary are underlined:

MIDSTSHSNLRSKAFIFGPQDLSFDVRSFNKLHSQLQNHQWVLDALASLPKLWD
NFAASDQKVQQSNTGKLLENLNAWISSGVAPEEAFPLPNVLLSPLVVIGQLVEY
MTFLKAAFPDLGKKHDLPISIKEDTETFGLCTGTLCAFAVACSSNIADIQHYGAV
AARLAMLVGAIVDTEEVLSDPEGKSVSFSAS<u>WNSAEFSDSFTHVLETFPD</u>AYVSV
IVDQRRATLTASKKTAPAIIERLKQEGAHVTSIALSGRFHWKKHQDAVSSLIQFCG
LDPDLQLADATKMLLPSRSSSDGQYITTGKLHELALRAILLEQSEWYKTCRISYLS
KFIMDDAAVICFGPERCMPPTLARKLGPRLTYVSEIDISSSRVPGQLLGGTQKLNL
TDLPDERIAVIGMACRLPGAEDHEGFWEILKTGQSQHREVPEDRFGMATAWREA
DKRKWYGNFIDNYDTFDHKFFKKSPREMASTDPQHRLMLQVAYQAVEQSGYFR
NNGTNRRIGCFMGVGNVDYEDNIACYPANAYSATGNLKSFLAGKISHHFGWTG
PSLTLDTACSSSSVAIHQACRSILSGECNGALAGGVNVITSPNWYHNLAGASFLSP
TGQCKPFDAKGDGYCRGEGVGAVFLKRLSSAIADGDQVFGVIASTKVYQNQNC
TAITVPNAISLSELFTDVVRQARLEPKDITLVEAHGTGTAVGDPAEYDGIRAVFGG
PIRSDVLSLGSVKGLVGHTECASGVVSLIKTLLMIQQGFIPPQASFSSINPSLNAKA
EEKIEISTRLKPWDAPFRAALINNYGASGSNASMVVTQPPNLTETPSTPLPGKSYP
FWISAFDQQSLQSYVRRLRQFLEKHAADKNLSVANLSFQVACQSNWSLPQALVF
SASTKEELNRALASFEKGSTDFPSVQLPDPKPVILCFGGQVSTYVGLDQEVYNST
AILRHYLDQCDAMCLSLGLQSIYPAIFQRSPIEDIVQLQTALFAMQYSCAKAWIDS
GLKVASVVGHSFGELIALCVSNAVSLKDAVKMISGRARLIKERWGADKGSMIAV
EADLSDVEALLAKVKSQMGSETGLAIACYNASKSFTLAGPTKDVDHAENLLKND
PDFSGIRYKRLNVTNAFHSVLVDALIDDLESLGQGIRFKEPTIKLERATEQESTSTL
NANYVATHMRKPVFFAQAVKRLSDKFPVAIWLEAGSNSTITAMASRALGTSNSS
FQAVNITSEGAFRFLCDTTVKLWKEGQKVSFWAHHRLQTPMYTPVLLPPYQFEK
SRHWMDLKVPPKPEASVQVAEQTAIIEAPKGLTTFVGYQDASQRSVRFRVNVTT
EKFNRLLSGHIMANAAAVCPGMFQVEIALDALTSLRPEFQARSFIPELHDLRHYQ
PLVRDESRAVWIEAHCPNAEGLVWNWKLTASDDKGSGSVTHTSGTITFQAADSV
QVKSEFEKLRRLIGRKRCLQLLDSNVADDILQGRNIYRAFTEVIDYKEIYRHVTKI
AGRDNESAGRVIKTYDGETWLDTVLTDCFCQVAGIFVNLMTTKIDLSERGIFICD
GIDRWLRAPNAGSNNTPSQVYEVFALHHCESDSKYLSDVFAFDAREGSLVEVAL
GISYQKVSISGIRRVLSKGMPAGLQPQVPTSPAAVAAIKTVSPPPVADSPLVDGSS
TAVSGTPPTKKAPKAPSVDITGKMREIICNLSGLEPDEVKDDSDLVELGIDSLMS
MELAREVDLAFKTTIDVTQLIDVTDFRSLVECMQRILGIDNQEDNTYLAEGLNGH
EGVVTNGNAYHVNGTNGVVNGNGVLFPELGGSILPKSAILDAFRIAKEATDDFIL
NGQLGTYYNEVMPRSTELCVAHIVNAFEQLGCPIRSAAAYQRLERVPYLPKHER
FMNLIYGLLEEARLIDINGSEITRTSVPVSTKSVETMLEELLHDEPLHAAEHKLTSL
TGSKFADCITGKEDGLQLIFGSPEGREIVTDVYAKSPINAVWIQQAEFFLEQLVKR
LPNTGEPLRILEMGAGTGGTTVKMLPLLERLGVPVEYTMTDLSSSLIAAARKRFK
KYPFMKFKVVNIESPPDPQLVHSQHIILATNCVHATRNLEISTRNIHRILRPDGFLL
LLEMTEQVPWVDFIFGLLEGWWLFEDGRRHALQPATHWKKILTSVGYGHVDWT
EGTRPEANIQRLIIALASEPRYDHTPQSLQPPVQVPLTDIAGRQEIIDTYIREYTEDF
RALPIPGIQQAVMPAPTGHCVLVTGATGSLGSHVVGYLSRLPNVHTVVCLNRRS

TVPATIRQEEALKVRGISLDDNSRSKLVVLEVETAKPLLGLPVETYQKLVNTATHI
VHSAWPMSLTRPIRGYESQFKVMQNLINLAREVAAWRPVPFKFSFQFISSIGVVG
YYPLRYGEIIAPEETMTADSVLPVGYAEAKLVCERMLDETLHQYPDRFRPMAVR
IAQIAGSTSNGHWNPVEHFAFLIKSSQTLKALPDFDGSLSWCPVDDVSATLGELLI
SNTTPYSIYHIENPSRQQWRKMVKTLAQSLDIPRDGIIPFDQWIERVRNSSASIND
NPARQLLEFFDQHFIRMSCGNLILDTTKTREHSATLRERGPVGPGLVEKYISAWK
TMGFLD

References

Geer, L.Y., Domrachev, M., Lipman, D.J., and Bryant, S.H. (2002). CDART: protein homology by domain architecture. Genome Res *12*, 1619-1623.

Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Honigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M.*, et al.* (2014). PredictProtein-an open resource for online prediction of protein structural and functional features. Nucleic Acids Research *42*, W337-W343.

Appendix B: Supplementary Material for Chapter 3.

Figure B1: Protein sequences for NR-PKSs used in this work.

See Appendix A for PksCT.

DtbA:

MEQPALVIFGPQISWPSNSSAERIRRDLVQEPSLAPFADAIRKIPELWASLVEIEPRLRETQGRRNLE
QLRDWVEIGSLPRRDSPDEILPNGFATPFTVIAQSIEYWRCFPRGGLQWLHSRSEAGPALEGMCTGF
LTAIAAATSKDSTEFFTMAAIAVRLAACIGTFVDLNGNTSDGTGEARALVVRWSSPIQESCLLRILE
EQPNAYVSVLKDETSRTITVPASEIPSTAEKLLMNGLRVHALELKGRFHHASNEDIANKLLQMCQL
NADLHFPDAQGLLTTLRRNSDGRRLSKGPLHIIAIRSLLLDQVDWLSTMTSALESAVIKDKPPKILV
FGSVNVVPHCLIQRSGARVCVSNDMIQPLSNITTTVKDFVPEPDMTKPGVKSLPDAPGCDNHDNSI
AIVGMACRFPGADSAEEFWELLECGKSMLSVLPESRFPTKGLRRSADDTVFWGNFIQDPDQFDNR
FFQISSREAASMDPQQRLILQVAYQALESSGYFDGQRKPTHTNIGCYLGVGSVDYEANIYSHQPNS
YSALGSLKAFVSGRVSHYFGWTGPSITYDTACSSSVVALHSACKALLSGECTSALSGGVNVITNPA
LYQNLRAASFLSPTGATKAFDAAADGYSRGEGCGLFVLKRLENALADGDQVLGVIAASAVNQTQ
REYGTSITAPVSKSQQSLYRHVLSQAGLEPSDVTYIEAHGTGTPRGDPIECESIRTVFGSTNTGRQQT
LYFGSVKGNIGHLEAGSGAAALFKSVLMLQKKKIPKHMNFNKLNPQIPPLEPEKLAIPTKTISWDA
PFRAICISNYGAAGSNGALVLREAPEPLKSTCARPTGNGPLRCQLISIAADSAESLQAYCNQLLRYL
ADVNERHFPDVAYKLARIQNPAHKYRLSFVTRSISELRSFLSAQSHVTPTPSFSGTVKPVVLCFGGQ
SKSSVGINKQLYDSISLFRHHLDRVEAACQALGKTIYPFIFSKEVIQDSVTLHCCLFASQYACARSWI
DAGIKVATVIGHSFGQISALCVCGVLSLEDAVKYVIERALLFDINWSAEKGTMLALKGDTQIVTRL
VSTSQVEIACYNGPRDVVVAGSIQAIEGFQSRAKAAGIDVRCLDVARAFHSSLMDPALPQLRRIAN
ALTFYRPYIPIQTCTQGHGRDSFDPEFLVDHSRQPVFFHDAVQRIESQLGPCTWIEAGSSSFGTTLAS
RIASQLNSFIPVDIGGPSPAESLANATLKLQALGYNVQFWAYHRSQRDAFRVFNVPPYQFAKSKH
WLEFKEFNSLAKESSHTDDNDQHSKLLLFVQSRDREKGIAEFKVNANSDYFRACVSGHAVLGHSL
CPASLYLELAAQAAVYICENESPVLPQVEDLQISAPLGLNPSATIKLLLSLKEPKVWRFSLLSTTGA
EDIHHASGTIQVPANADFLVSESSLYERLIGADRCNSILQDASIKSINGLVYNVFGAVVDYKDFYQG
VQRISATKSEASGVVSLPVAAAGIMQDSVYDPLALDNFLQVSGIHINCFRDFGRQEVYVCTSIGKL
KVYQSLEKVRDQSWLVYSNTKQGGPGVLESDIFAFNKTTGSLMVALFGVRFAKVSVNSLTKTLSA
LNHGKPVQRLKDSTPPSTPLSNLGLGTSPKTELDGNVLLRLQRLLSRVTDCPVEDILQVNSLEALGI
DSLMRSEVTAEIRQEFGLDISTDTLAKASTLFHLTTLISSSQKWPVQIVNTPSSSESHTTLSMRDSTSS
TLKDILSAITDIPVDEIGNDSTLEALGIDSLMRKEVQSELRKNLGRADIPENLHEYSVTSLAAYLAD
QDCGKRNDLSCGTPPGSVSLGRHCDIKALGSHTQLMAPEMISGLEERQELVKRNDRSQLFQLSSPP
KPLCSSTVGGFTLKDLAASFEEIKGAFDQLSDESQFLNFYAQVHPRQMKLATAYVVEAFRSLGCPL
SSMSPGESLSLSPVSYEPRHGNLMKQLYRLLQDAGLISLDNLGEYKRTSKPVDCTSSADLLDAIIQD
FPQHQLEHRLLSITGSRLASCLSGELDALDLIFGDRTAKDLVSDVYLKAPVFVTGTKLLCKFLLSCL
RQRTQPIRILEIGAGTGGTTADVVRCLISSNIPFEYCFTDLSPSLVAAARKKFKWCPNMEFNVLDIEQ
EPPSVNQYDLVVSTNCIHATRNLTATTKNIHKLLKEGGILCLVELTRNLPWFDLVFGLLEGWWLFN
DGRQHALAHELFWKDSLRQAGFSYINWTTGDTKESEQLRLIISVKMDHQYDLSKWETGSEKCTG
DNFEIVTSRQPLSVYSTPDLLKEPGLSGSVVLLTGGTGNLGTHVLHQLINRADVRRVICLNRLTTND
DPIQRQRRALRDKGIDLKERQWEKIEVLEAKSSHTALGLQTEQYQRLRDQVTHIVHNAWPMSFKR
SLHTFEPQFKTLQNLLKLCHEAKYGARLLFISSIGVVGRHPNTFANKPVPEDPVRDCQSSLGFGYSQ
AKYHCEQIINRALEQDTRLEASYVRIGQITGSQHFGLWNTEEHVPALLRTSQTIGALPHLDGLASW
LPIDIAAATVSELLLDTAHLRMVYHVENPVRQPWCELLGYLSAHLRLPIIPYKEWLSRMETNGDSV
TGSPNPAKNLRDFFKNDFLHMSCGSVVMSTTSTASVSAALRSAGPVSPGTLLLYIEQWRRTGFLG*

MpaC:

MNFHKGQPKEDLRVLFGPQCPDITDSITHIRDAISKDPTGLGFLTNILDELPSLWPTIAGAWPALKN
VEGESQLLALGRLFEHESEDRVEASNLMMTPITVMRHIVDFWNLQDVATHPAFPSSSLSETEMPRI
VDTQGFCVGLLAAIAVACSRNTQEFQYVASNAIRLSLCVGALVDLDEILCGSTTSLAVSAERVKQE
IHDHGLRTKQLSLRGRFHHEAHREGIQHIMKLCTNDSRFKLPRSDALLTPLRSSQGGEIFQQEALLH
TVALDSILCAKANWYDVVSALINSTEMTVDQSHLLSIGPEEFVPRSARSRSVARRELQSYAMQGFS
NESPQPSTASLSNSVQTFDSRPQAAEASPIAITGMACRYPNADTLAQLWDLLELGRCTVKSPPESRF
HMSDLQREPKGPFWGHFLERPDVFDHRFFNISAREAESMDPQQRVALQVAYEAMESAGYLGWQP
NGLSRDIGCYVGVGSEDYTENVASRNANAFSITGTLQSFIAGRISHHFGWSGPSISLDTACSSAAVA
IHLACKALQTNDCKIALAGGVNVLTNPRVYQNLSAASFLSPSGACKPFDASADGYCRGEGAGLFV
LRPLQDAIDNGDPILGVIAGSAVNQGSNNSPITVPDAEAQRSLYNKAMSLAGVSPDEVTYVEAHGT
GTQVGDPIELDSLRRTFGGPQRRNSLHIGSIKGNIGHTETSSGAAGLLKTILMLQQQRIPRQANFNQ
LNPKVKSLTPDRLVIASESTEWASTERVAMVSNYGASGSNAALIVKEHAPIRSEQNGTAPEYIQNV
PILVSARSEESLRAYCGALRATLLSHPPSETLVQKLAYNLAMKQNRDLPLNLTFSTSSDATSLSARL
EAISTGASADLIQKRPSNEPPVVLCFGGQNGLTATISKEVFDASALLRTHLEDCEEVGRTLGLPSLFP
TIFSSAPITNIIHLHFILFSIQYASAKAWLDSGLRVSRIVGHSFGQLTALSVAGSLSVRDGIHLVTERA
RLIESSWGPESGIMLAVEGTDIEVQQLLDQTGHIADVACYNGPRQQVLAGTAESIAAIENAAARTP
SASKLRLTRLQNSHAFHSRLVDSIVPAIMEVAGSLVYQTPIIPIEACSASGDWSTITAAEIVEHSRMP
VYFRRAVERVAEKLQAPAVWLEAGSASPIIPMVRRVLESSSVANTYHKIDLGGSSGAQNLANVTS
ALWAQGVHVQFWPFDRAQHGSFKWMNLPPYQFAQNSHWVDFDPAAFSSAGPSSGKQSAGQEAG
LLCQLSESPDERLYHVNIQDALYRACTQGHAVLNQTLCPASMYMEMVLRAAASIFPTGNASEPAM
SHIEDLTISSPLVLDPQGDVFLRLTSDGAGPTRPWLFSIFSSESNDHTSVHAEGTVCLHQERSRALAR
FQSMDRLLDSARSKTIEADPASNGLKGSTVYAALESVTNYGDYFRGVKKVFANGREASGLVSMM
PSASETNCDPILLDNFLQVAGIHVNCLSDRRSSEVFVCNAIGETFVINSLLKQKNGASPSTWKVYTS
YVRPSKTEIACDIYVMDCQTDTLSAAMMGVRFTSVSIRSLTRALAKLNNNVLETAEAQSVVEPAIP
AEKSVVTATPSAPAADGGGAKDLATVQEMLCELFGVSVAEVSPSVSLVDIGVDSLMSTEVLSEIK
KRFQVDMSYTTLVDIPNIQGLVEHIFPGHSHAAPSQPVVETAPVQSVAPQAVSHVPTPANNGPPLV
SVARQCFDTTHAAVSHTSDAHWTGFFHTTYPKQMTLLTAYILEAFRALGSPLEASEPNEVLIPISVL
PRHEQLRKHLYKILESVGLVRQMPTGELVRTTTPIPLSQSHDLHTQIRAEYPPYALEHDLLQITAPR
LADCLTGKADGVSLIFQDANTRRLVGDVYAQSPVFKSGNLYLARYLLDVVQSFGSSRTIKILEIGA
GTGGTTKNLLEKLSTIPGLSTRLEYTFTDISPSLVAAGRKTFANYNFMRYETLNVENDPPSALSGQY
DIVLSTNCVHATRNLRESCTNIRKLLRPDGILCLVELTRDIFWLDLVFGLLEGWWRFEDGREHALA
TEMMWDQTLRQSGFEWVDWTNNETVESNALRVIVASPTGNSSTATMSPSKLTKMETVVWGERD
NLQLRADIYYPETVDTTRKQRPIALMIHGGGHVMLSRKDIRPAQTQTLLDAGFLPVSIDYRLCPEV
SLAEGPMADARDALSWVRRVLPNIPLLRADIRPDGNQVVAIGWSTGGHLAMTLPFTAPAAGISAP
NAVLAFYCPTNYEDPFWSNPNFPFGQTVASNEMEYDVWEGLQSMPIAGYNPALKERPLGGWMST
RDPRSRIALHMNWTGQTLPVLLKACTIKGNTEKCSPDDLSRPTEEEIQAVSPNYQIRVGRYNTPTFL
IHGTSDDLVPCAQTESTHGALTASGVEAELRVVQEAAHLFDLYPASHAGQEAKAAVAEGYEFLRR
HVQL

PkeA:

MLGHRDFTTLPLSRREFLLFGPLALSFDQAAFEHLRKTIVNSEEHRWALEVLGSLPQYYATIVNAFP
GINGRNEVQLEDLKGALHSGKPLATSFPLPNTLLIPLVMVLHLTEYSRFLQEISEELESGIDLFDASR
HNKETVGFCTGLLSAMAVSSAGSREDFRKYAAVAVRLGLLVGVVVDSHDISSAQGPSKSISASWN
SAQKREDARRIMDEFPQAYISVYYDEDRATITAPASEISDLHRRLRASGIVTAEIGLNGCFHADCYL
DQLDPIIQFCDSQPDFQLPDASKVVIPTRSNATGELIRDGALHQHALRSILVEPPQWFESFTAVRDA
CAEDEGAIIFSFGPERCVPPSLLRVLSQKVVTVEDLDVLKRYQYSYSENDIAVVGMSCKVAGANNL
EEFWDLLCTGKSQHREVPKERFSFETVFRDVDSKRKWFGNFIDGHDQFDHKFFKKSPRESATMDP
QQRHLLQIAYQAVEQSGYFHSANPDRQIGCYMGVCACDYENNIACHAPNAFSATGNLQGFIAGK
VSHFFGWTGPGLTIDTACSSSAVAVHQACKAIITGECTAALAGGTHVMTNPLWFQNLAGASFLST
TGQCKPFDAKADGYCRGEGIATVFLKKLSAAVADGDQILGVITATAVQQNQNCTPIFVPNVPSLSD
LFRVVVKQSRLQPSDVTVVEAHGTGTAVGDPAEYDSIRSVLGGSSREKTLALSSVKGLVGHIECTS
GIVSLIKVLLMLQKRMIPPQASFTTINPAIKATPADKINIPTTVKTWDAEFCAALINNYGASGSNASI
VVTQPPVGTVKPSAETSGLKYPFRFCGMDEQSLRRYSKIFRQFLNRKSYSAQDLSLRNISFNVNRQ
SNRQLDRTLLFSVKTLEELEQKLVTFENDNDSITSLALPKSKPVVLCFGGQVSTFVGLDRTVYERV
AILRKHLHTVDAVARSIGLKSIFPRIFETTPVSDTVHLQIMLFASQYACARSWIDSGIQPVAVVGHSF
GELTSLCVSQSLSLEDAVKMIAARATLIRDAWGPEKGAMLAVEADLEDVQKLLAESSAGCQDVQ
PATIACYNGPRSFTLAGAVAAIDAVAEALATPAFSSMKNKRLNVTNAFHCALVDPLLDRLEESAR
ELTFRAPVIPVQRATEYQTEELPTSRFVADHIRSPVFFNHAIHRLADKYPSCVFLEAGSNSTVTNMA
SRALGNPSSSHFQAINITSHNGWNNLVDATMNMWKSGLGVHFWAHQPSQTKEYALLLLPPYQFE
PSRHWIELKNPPKLTAAPAIEEVKKEEAKVPNTLLTFVGYQDSERQQARFRVNTMIPKYDKLIRGH
IIAQTAPICPATVQLDLVIESIRSIRPELASTEHEPQIHAVENLAPICVNPLRAVWVEVTADDVAQGT
SWNFQVYSDDLQNGFSKTIHTTGRVIFRSISDVSLKYEFARFERHFRHQTCVELMRGGEVDEVLQN
RNIYKMFAEIVDYGEDYRGLQKLVSKGNQSAGYVVKKYNPESWLDGHLADSFCQVGGIYVNCM
TDRVPNDMFIANGIEQWMRSPKMRQQDPRPESYHVLATHHRPSDKAFLTDVFAFDSTTGVLIEVIL
GISYVKIPKASMSKLLSRLTVNDSASCPTNMPLLSKSASVNLFDAPENLSTPSLSVAPTQQSAPALS
LSKVKKVKNDGPDKGQLTQRIKSILAELSGLEIAEIKDDSELADLGIDSLMGMEMAHEIEKAFTISL
PESDLMEVVDVPSLIKCVRKAMSGDADSAEYTTEQSTSEAADSDDKSTNYTTPSTPGEEALDMDK
SMREFLGKEGTELNLPFETVMKAFNETKNMTDDRIAEYQQTRYVESVLPMQSQMCVSLVLEAFD
QLNMRIRTAPAGEKFTRISHPKEHTRLVDYLYKMLEDASLINIDGEVITRTAIQVPRPSKEIFDELVS
QHPDQNAADKLTFYTGSHLAEVLKGETDGIKLIFGTQDGRELVSKLYRDWPLNRLFYRQMEDFLE
RLTSKLDISQGVIKILEMGAGTGGTTKWLVPLLAKLNIPVEYTFTDIAPSFVAAARKKFSKQYPFM
KFRTHDIEKAPADDLIGSQHVIIASNAVHATHSLSESGKNIRKALRPDGVLLMLEMTGTLHWVDIIF
GLFEGWWYFDDGRTHAVTHESRWAKDLQAVGYGHVDWTDGVRPENKLEKLIIAFASGGRYERL
HIPRPLESASADCAARQAVVDRYVQEMTAGFGAATGVSPSAPLAHQEPKGCCVLVTGATGSLGC
HLLAALTSLPTIASVVCLNRRSRQDPLERQHRSLLEKKIFLSEETAARVRVIETDMSKPQLGLLEEE
YNYLLNSVTHIVHNAWLMNAKLPLRRFEPQLQIMRNLLDLAYGISLQRPMEKVSFQFISSIATVGH
WPIWTGKSSVPEERMAIESVLPTGYGDAKYICERMIDETLHKYPDRFRAMVVRPGQVAGSSTSGY
WNTMEHFSFLVKSSQTLNALPDFDGVLSWTPVDVVASTLVDLLLLPEDKTPYSIYHIDNPVRQPW
KEMNVVLADALHIPRSNIIPFEKWIQRVKDYPRQVEGAEGDNPAILLVDFLDNNFIRMSCGGLLLE
TKKSREHSKTLANLGPVSAETARLFIKSWIDMGFLSP

AusA:

MGSLDDNTLQQVSVLFGPKYPEVELPAGHIRRYLSNQRNANWLHDAIRDLPSVWHDILRLWPAA
EKLHGDARLRQLSAFLGGGTLRPDMAEPMNFLLVPATVLRHLVDFLELKEDKNYDVCDIQGFCV
GFLAAIAAACWSDNEDEFGKVVSTVLRLAVYIGAAVDLDELCEQPARSIAVRWRTAQEHKLLTEV
LTRYQGAYISCVTDENAVTVTVWDSQSVSFAKELEKHGLSVKTTTLRGRFHHSNHTQAVEDILQS
CERNSRLCLPSKCHKRSLPRSNINGRVCEADSLFTVAVESILTTQANWKITVTATLDNMGQSDARSI
IPIGAGQFVPRHARCRMLNIVEFNKGEHINGRRKMQSATALDVGVNVTAPETTAVPIAVTGMACR
YPQADSVEELWRILDLGQCTVSPMPNSRLKSGSLQREPKGPFFGNYLARPDAFDHRFFGISAREAE
SMDPQQRVLLQVAYEAMESAGYCGLRRSKLPDDIGCYVGVGCDDYSENVGSRNATAFSATGTLQ
AFNSGRISHYFGWSGPSVTVDTACSSAAVAIHLACQAIRTNDCAIAVAGGVNIMTDPRWSQNLAG
ASFLSPTGASKAFDADANGYCRGEGAGLLVLRPLEAALRDGDPIHAVITGTSVNQGANCSPITVPD
SNSQRSLYLKALSLSGLTPDVVGYVEAHGTGTQVGDPIEFESIRKTFSGPNRATKLYVGSIKDNIGH
TETSSGVAGMLKTILMIQKRRIPKQANFRRLNPRITLNERNHIEIPTQSIDWEAEKRVAMVTNYGAA
GSNAAIVLREPASTPATSNSAHRETLPSHVPFYVSARTEESLRSYCEALQSTIREVAQSGTNTVQHI
AYNLARKQNRDMEHFVTFPAAAGEPSELMTRLGSIASAHTQVERRSQSFHPVIICFGGQTGDTASIS
RNLFESCELLRFHVVSLSFSFWIRDECENACNALDLPSLFPAIVSPFPNKDIVNLHCVLFSIQYATAK
AWLDSGLQVTRMIGHSFGQLTALCVAGGLSLIDGMRLVATRAQLIQKHWGPHTGVMLSLRASKE
KVQALLDAASGHADLACLNGPDNFVVAGDEESIRRIEIIATEKGMHVELKRLKNTHAFHSRLVDAI
LPGLSEVANTLTFRQLDIPVEACAEQEDDWLWVTGDKIVQHSRKPVFFHDAVERTLSRVDGPCVW
LEAGTASPVINMVRRVVEASRPLKSHVYLPTDLSGAQAQANLAKVTCTLWSKAVPVQFWPFHPSE
TGYRWINLPPYQFAKTSHWIEYNPDAFRSPPQVPDQENVQEASLVRLLRQDGKEALFTINNKDNV
FRMCTAGHAVANQNLCPASLYFELVVQAALLVSSTATKPTMYHIESLNICSPLVLGMPGAVLLQL
TQQDESHGQWSFVLSTRDGLQDAVTHATGRVSLQAAGSNTGICARLSSLQRLLNLASWNSIATSP
SSSGLKRSTVYQAFARAVNYADYYRGVEEVYAVGHEATGRVILPSSPTKCNPCDPILIDNFIQVAGI
HVNCLSETHDDEVFVCSSVGDVLIGESFVRRDTAATVPWAVYSNYEPESKKKIVCDVFVLDHTTG
ALAVCMLSATFTGVSIQSLKRTLNRLSNHTARPTEAEQVSINVAAEATALSSTPVAHVSSSDGDLL
AVQTMLGELLGISADELSAAAALGDIGVDSLMSTEVLTEINKRFGVAISNAELTQIPDVGGLVQRIF
PGHSVVRIKTHSQGAVETEITITDREPKSISVDLAPVCDTSPTAFVDKASKLFATTRTSAEFSRKTRF
AGFCDTVFPQQMELVTSYVVEAFHALGADLASLTPGQVVPPVKILPQHGKVMNQLVAVLEYSDLI
ERRESEIIRSQQPVGTVPSLILYKKILNKHAQHASEHKLLHTTGSRLAECLSGKADPLSLLFQNAEA
RALMTDVYSNAPMFKSATIQLAQYLKDLLFNLGTQREIKVLEIGAGTGGTTNYLVQELAAVPGLR
FQYTFTDISSSLVTLARKRFKAYDFMRYTTLDIENDPSPELQGQYDIIISTNCIHATRNLITSCTNIRR
LLRPEGILCLIELTRNLFWFDLVFGLLEGWWLFNDGRSHALAHERLWDHNLRQAGFNWVDWTDN
DSAESDILRLIVASSTQPFYALEGDDECEADCNTVQEQTVLYNTRDGLELFADIYYPEKTDRSGAK
RPIALLIHGGGHIMLSRKEIHHEQVRMLFDMGFLPVSIDYRLCPEVSLLDGPMQDACDALAWARN
KLPQLQLQRRDILPDGNNVVAVGWSTGGHLAMTLAWTAPARGVSAPEAILSFYSPTDYTDPFWS
KPNFPYRVDVSTSDIQTGNPLDALQDAPISGYNPPPSKRALGGWMAPSDPRSRIALYMNWTGQTL
PVLFYGCNYRARAAESGQDYEVVLPEPILSEVQKVCPFSQISAGSYRAPTFLIHGTLDDLIPVQQAQ
RTHDKMQACGVDSDLRIVRDGLHLFDLEANFAGNQHAFQAVVDGYEFLRRHVGL

PkbA:

MAPAPSALVFGSQTTLPSVEAASRLRAALLLDPRLYRMRTSIESLPEIWPALAISDPALERVAGPAE
KSLRQLCRWLSHSEFPDATEISELPATFVTPFTVILHTVLYMHYTDENGSRGHADVLRAVRNNGG
VQGFCTGFLTAVSVATSPDLEALSRQASVALRLAVAIGAYIDLDLLDSDVSSVAVRSRAGKKGLEE
TLAQFPGAYISVITDELNATVTAPRASLDALSQSLASNGLSAKRFDLRGRFHHPAHQKALEGLYNL
VASNPVFQFSHSELLAPVYSNIDGQLLSSDSIIDTLLQSILVQRCDWYASISTALHKRSNGEKTSVVQ
FSLVECIPPSVLRQARLSVQRITDVPVQNSPATVPAPLNAVRARIQTSEPIAIIGMGCKFPGADTLDE
YWQLLAQGTSMCRTMPEERFKTSSLRRSPGEKLKFWGNFVNDVDAFDHRFFKKSSREAASMDPQ
QRLVLQVAYQTLESAEYSGLSGIKASRDVGCYLGLCASDYTDNVASHPPNAFSSLGTLRAFLSGKI
SHFFGWTGPSITYDTACSSSAVAIQAACRALQTGECSMALAGGVSLYTSPNFYQNLSAASFLSPTG
PTKPFDAKGDGYCRGEGVGLVFLKPLSSALADHDNIMGVIAAAAVNQNQNSTAITVPHSESQIELY
RKVVSEAGLHPHDVSFVEAHGTGTPVGDPIEFTSIRTVFGGSNRANPLAIASVKGNIGHTEGASGV
ASLIKTVLMLQRGLIPMQANHTTLNPKIAPLEQDKVIIPHTTQPWSANFKAACINNYGAAGSNAAM
IVTEAPTGARSEKQGTLPKCPIYVSANTVSSLKEYCKELLRSLRGRSPDCLASLAFQLANSQNRGFP
HALITSVTSKAELEDQLSAVVENRNNSLHTVAPTERPVVLAFGGQVARSVGLSRQVYDSSAVLRT
HLSNCDDILTSAGLNSIFPAIFRKEPIADVVLLHSALFSAQYACAMTWIEAGVIPAALIGHSFGQLTA
LSVGRVLSLKDGLGLITERAKLMRDAWGPEHGSMVSVQADVQTVARLMKAAETKDPKDALEIA
CFNGPTSHVVVGSADAADRLEAALTQESIRYKRLAVSHGFHSKFTEPLLLGLEQCAERLTFRTPKY
AIETCSSGSSWSEFNASMIVQHTRTPVYYTEALARIEAKLGACTWLEVGTGGSVAGMIRGALNVPS
DHLIQAVNLAGETGTAALADATVNLWKSSHKLQFWAFHRSERECYQPLELPPYQFEKTRHWLDW
KDTMTEQATVTQSTREETSVEEFLTFVKYKDSTKQQAEFRISTEHEKYSFFVKGHAVLAEPLCPAP
LYIDLACKAGQMVYSDTSETLIIPSVEDLEIQAPLGVGDRVIILRLQQSPFLKTAWTFCFCSRPVMG
NSAEEQLHASGTVVLRENDTKTAAEFSRFGRLVSSKRVQEMKSDPDCHILQGPVVYQLFSRVVSY
ADYYKGVQSVYASGAEVTGRIRLPPTVKDADTRRPLLVDNFIQTAGIHVNCLTDVGAKEVYVCTK
VDRVQSAAAFTEDLANVDASWIVHSSYHPTSEKEVVNDIFVFNAATGELAMFILGAHFTRVQISSL
GRVLSRANTADAAPIKVAVPVQSPALRAQPKRVLLPPLTKRSITRPTLEISEKLKKTLSRVVEVPVA
DIHDGGILADLGVDSLLGTEVLTEINQVFNVSIPADEFALLTDVASISKCLASYLGVHDSGSQPEDL
ADADSVESDSDMPTGAVTSGITTPDDAVSRLADLLAENLEYDGTIEASNNLADLGLDSILSIELAND
IKKIFNCDVDMSQLNMESTFADLIALVPALNIEQSLSSVPASLTTQGSDFEMAQHAFEQIRFDYDIY
TKETGFYDFWKRVYPAQSRLVLAYTVEAFAQLGCDLALMHPGDRLPKIGYLPAHEKLVQQLYNI
LRDGMLVATSDSGFVRSDKPVDPTSSTRLLEEINTIAPQHASEHSLLHITGSKLADCLTGTADPLNL
LFRSKANRDLLAEVYLNGPMYAAISRLLCSFLGNAFSDRQSSGTFQILELGGGTGGTTGHVLDYLV
RSGIPFTYTFSDVSGSFVAAARKKFAGRPYMEYQVIDIEKEPADSLTGKFHAVISTNCIHATTNLEIS
TGNIHRMLRPDGFVALVEFTRNMYWFDLVFGLLEGWWLFEDGRQHVIASESFWNTSMRKAGFQ
HVSWTDGDSFEARTLRIIAGFRAPAVNEIYTPRLDSRDTETAVESVMYKQADGIPLFADIYYPPSVS
NEPRPIALMIHGGGHIMLSRKDIRPKQTAHLHTLGFLPVSIDYRLCPETTLTEGPMRDVSDAMAWA
RSTLCSLPLLCPGLTLDPSRVVVVGWSTGGHLAMTTAFTSIERGLSPPDAILAFYCPTDYEDRVWT
QPNYPENTDVDGLSLMKYNLLEGVQDRPITAYNIPLTSRSNSKAGGTPAGGWMAPDDPRSRIVLH
MNWKGQCLPVLLRGLPPSNSLSPGDAEKLISQPQPEIEEIQRVSPYAQIRRGVYRTPTFVIHGTDDD
LIPYEQSVRTVQALKDMGVRAEVSVPQGKAHLFDMFKDADGSSWEVVKRGYDFLKEEVSGK*

Table B1: Primers used in this work.

See Appendix A for PksCT primers

Primer Name	Sequence
AnAusA-CMeT-5	GCATCATATGTCGGTGGATCTTGCGCCAGTATG
AnAusA-CMeT-3	GCATGCGGCCGCTAGAGGTTGAGTAGACGACGCAAC
AnPkeA-CMeT-5	GCATCATATGCGCGAGTTTCTAGGGAAAGAG
AnPkeA-CMeT-3	GCATGCGGCCGCACTTTCTAGAGGTCGGGGAATGTG
MT-V1847.5	GCATCATATGGTGAAAAGAAATGATAGAAGCCAG
MT-SS2239.3.2	GCATGCGGCCGCGCTTATTATCAAGCGGAGCTGCTC
DtbA-ACP-5	GCATCATATGGGCACATCACCAAAAACAGAGC
DtbA-ACP-3	GCATGCGGCCGCCGAAAGGTCATTCCTTTTACCAC
AusA-ACP-5	GCATCATATGTCTACCCCAGTGGCTCACGTTTCG
AusA-ACP-3	GCATGCGGCCGCGATACTCTTTGGCTCTCGGTCTGTG
PkeA-ACP-5	GCATCATATGAGCCTCTCCAAAGTAAAAAAGGTC
PkeA-ACP-3	GCATGCGGCCGCAGACTTGTCCATGTCGAGAGCTTCC
AnPkbA-CMeT-5	GCATATTAATGAGCCTGTCGTCTGTTCCGGCG
AnPkbA-CMeT-3.2	GCATGAGCTCGAATCGAGTCTTGGTGTATAGATCTCGTTGAC
AnPkeA-ex1-5	CATGCATATGCTTGGTCATCGGGACTTCACTA
AnPkeA-SKM1-3	GCATGCGGCCGCATTCGGTACCTTAGCCTCTTCTTTT
PbMpaC-ex1-5	GCATATTAATGAATTTCCACAAAGGGCAACC
PbMpaC-ex1-3	CTCCTGCTTCACTCGTTCTGCGCTGACTGCCAGTGACGTGGTTG
PbMpaC-ex2-5	CAACCACGTCACTGGCAGTCAGCGCAGAACGAGTGAAGCAGGAG
PbMpaC-ex2-3	GCAATCATTTGTTTGCAGAGCTTTGCATGCCAGATGAATAGC
PbMpaC-ex3-5	GCTATTCATCTGGCATGCAAAGCTCTGCAAACAAATGATTGCAAGATC
PbMpaC-ex3-3	GGGATCCCCGACTTGGGTGCCAGTGCCATGGGCTTCC
PbMpaC-ex4-5	GGAAGCCCATGGCACTGGCACCCAAGTCGGGGATCCC
PbMpaC-ex4-3	CCTCCGTGGATCATCAGGGCAATCGGTCGCTGCTTCCGAG
PbMpaC-ex5-5	CTCGGAAGCAGCGACCGATTGCCCTGATGATCCACGGAGG
PbMpaC-ex5-3	GCATGCGGCCGCCAAGCTGAACATGTCTTCTCAAGAACTCATA
PbMpaC-CMeT-5	GCATATTAATGGTATCCCATGTACCGACACCGG
PbMpaC-CMeT-3	GCATGCGGCCGCGTCTCGCTCACCCCAAACC
PbMpaC-SKM-3	GCATGCGGCCGCAGACGAAGGTCCAGCAGATGAAAAC
PbMpaC-ACP-5	GCATATTAATGGCGCCGGCTGCCGATGGA
PbMpaC-ACP-3	GCATGCGGCCGCTGGTGGTCCATTGTTAGCCGG
PbMpaC-SKM-3.2	GCATGCGGCCGCTCCTGCGGACTGCTTTCCAGACGA
PbMpaC-CMeT-3.2	GCATGCGGCCGCTCCTGTCGGCGAAGCGACAATGA
PbMpaC-CMeT-3.3	GCATGCGGCCGCTGTCGCGGTGGATGAGTTTCC
AnAusA-ex1-5	CATGCATATGGGGTCCCTTGATGATAATACTTTGC
AnAusA-ex1-3	CAGTCACGCATGAAATGTATGCACCTTGATAACGAGTTAAGACTTCC
AnAusA-ex2-5	GGAAGTCTTAACTCGTTATCAAGGTGCATACATTTCATGCGTGACTG
AnAusA-ex2-3	GCAATCATTCGTTCGAATGGCCTGGCACGCGAGGTGAATAG
AnAusA-ex3-5	CTATTCACCTCGCGTGCCAGGCCATTCGAACGAATGATTGC
AnAusA-ex3-3	GGGTCTCCGACTTGGGTGCCAGTGCCATGTGCTTCAACATAGC
AnAusA-ex4-5	GCTATGTTGAAGCACATGGCACTGGCACCCAAGTCGGAGACCC
AnAusA-ex4-3	GTTGCAAGCGTTTTCACATTCATCGCGTATCCAGAATGAAAAAGATAGAC
AnAusA-ex5-5	GTCTATCTTTTTCATTCTGGATACGCGATGAATGTGAAAACGCTTGCAAC

AnAusA-ex5-3.2	GCCGCCGTGGATGAGTAGGGCGATTGGTCTTTTTGCTCCG
AnAusA-ex6-5	CGGAGCAAAAAGACCAATCGCCCTACTCATCCACGGCGGC
AnAusA-ex6-3	GCATGCGGCCGCTAAGCCAACGTGTCTTCGCAGGAACT
AnAusA-SKM1-3	GCATGCGGCCGCACTCGCTTCTTGTACATTTTCTTGATCC
AnPkeA-ex1-3	CGTCATAATAGACAGAGATGTACGCCTGGGGAAATTCATCCATGATAC
AnPkeA-ex2-5	GTATCATGGATGAATTTCCCCAGGCGTACATCTCTGTCTATTATGACG
AnPkbA-ex1-5	CATGATTAATGGCTCCGGCACCATCCGCTCTTG
AnPkbA-ex1-3	AATTCATCCGTAATTACTGAGATATATGCACCAGGGAATTGCGCGAGA
AnPkbA-ex2-5	TCTCGCGCAATTCCCTGGTGCATATATCTCAGTAATTACGGATGAATT
AnPkbA-ex2-3.2	CTGCACCATAGTTATTGATGCAAGCCGCTTTAAAGTTCGCAGACCAG
AnPkbA-ex3-5.2	CTGGTCTGCGAACTTTAAAGCGGCTTGCATCAATAACTATGGTGCAG
AnPkbA-ex3-3	GCCTCCGTGGATCATCAGGGCTATCGGACGCGGTTCATTAG
AnPkbA-ex4-5	CTAATGAACCGCGTCCGATAGCCCTGATGATCCACGGAGGC
AnPkbA-ex4-3.2	GCATGAGCTCGACTTCCCAGACACCTCCTCCTTCAAAAA

Table B2: Plasmids used in this work:

See Appendix A for PksCT plasmids

Domain	Sequence	Plasmid name	Parent vector	Tag	MW (kDa)
PkeA SAT-KS-MAT	M1-N1286	pEAnPkeA-SKM	pET-24a	C-His6	142.2
PkeA ACP	S1650-S1783	pEAnPkeA-ACP	pET-24a	C-His6	15.7
PkeA CMeT	M1784-S2183	pEAnPkeA-CMeT	pET-24a	C-His6	46.8
AusA ACP	S1626-I1737	pEAnAusA-ACP	pET-24a	C-His6	13.0
AusA CMeT	S1738-P2123	pEAnAusA-CMeT	pET-24a	C-His6	44.6
DtbA ACP_2	G1619-S1853	pEDtbA-ACP	pET-24a	C-His6	27.0
DtbA CMeT	V1847-S2239	pEDtbA-CMeT	pET-24a	C-His6	45.2
PkbA CMeT	S1818-D2152	pEAnPkbA-CMeT	pET-24a	C-His6	38.4

Figure B2: UV-Vis spectra for compounds 1-12.

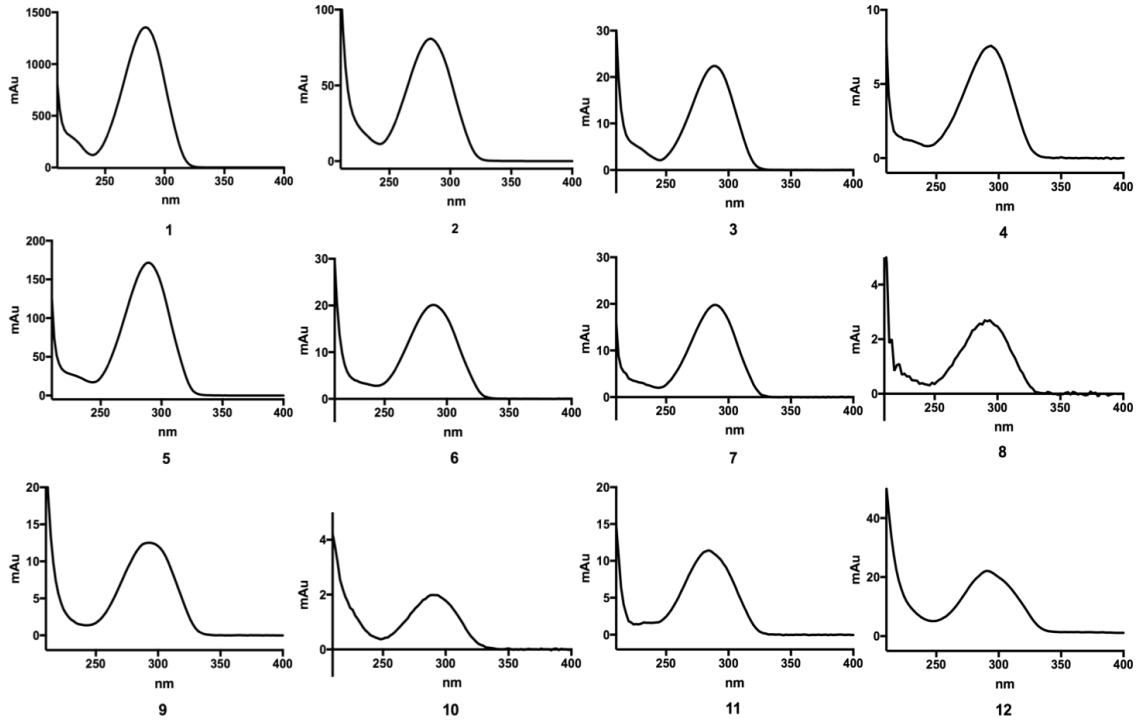


Table B3: Detected masses for compounds 1-12.

#	Calculated [MH+] or [MNa+] m/z	Detected m/z (ppm)	Formula
1	127.0395	127.0393 (1.57)	$C_6H_7O_3$
2	169.0500	169.0496 (2.37)	C8H9O4
	127.0395	127.0389 (4.72)	C6H7O3[-C_2H_2O]
3	141.0552	141.0553 (0.71)	$C_7H_9O_3$
4	155.0708	155.0704 (2.58)	$C_8H_{11}O_3$
5	141.0552	141.0553 (0.71)	$C_7H_9O_3$
6	183.0657	183.0650 (3.82)	$C_9H_{11}O_4$
	141.0552	141.0548 (2.84)	$C_7H_9O_3$[-C_2H_2O]
7*	197.0814	197.0804 (5.07)	$C_{10}H_{13}O_4$
	219.0633	219.0621 (5.48)	$C_{10}H_{12}O_4Na$
	155.0708	155.0701 (4.51)	$C_8H_{11}O_3$[-C_2H_2O]
8*	197.0814	197.0804 (5.07)	$C_{10}H_{13}O_4$
	219.0633	219.0621 (5.48)	$C_{10}H_{12}O_4Na$
	155.0708	155.0701 (4.51)	$C_8H_{11}O_3$[-C_2H_2O]
9	211.0970	211.0970 (0.00)	$C_{11}H_{15}O_4$
	233.0787	233.0790 (1.29)	$C_{11}H_{14}O_4Na$
	169.0865	169.0859 (2.37)	$C_9H_{13}O_3$[-C_2H_2O]
10	253.1075	253.1076 (0.40)	$C_{13}H_{17}O_5$
	275.0895	275.0893 (0.73)	$C_{13}H_{16}O_5Na$
	211.0970	211.0965 (2.73)	$C_{11}H_{15}O_4$[-C_2H_2O]
11	183.0657	183.0657 (0)	$C_9H_{11}O_4$
	141.0552	141.0551 (0.71)	$C_7H_9O_3$[-C_2H_2O]
12	141.0552	141.0552 (0)	$C_7H_9O_3$

* 7 and 8 co-elute by UPLC-ESI-MS.

Appendix C: Supplementary Material for Chapter C.

Table C1: Primers used in this work

Primer Name	Sequence
MpOrf4-g5	ATGAAAGGGCAGACAGGGCTTCGC
MpOrf4-g3	CTAGGGAGCACCCGTCTGCGTTGC
MpOrf4-5.2	GCATCATATGGTCCAGACGAATTTAGAGGTGGTCG
MpOrf4-3	GCATGCGGCCGCGGGAGCACCCGTCTGCGTTGC
MpOrf4v2-S122A-5	GCTTGGATTTGCACAAGGCGC
MpOrf4v2-S122A-3	GCGCCTTGTGCAAATCCAAGC
MpOrf4v2-R36A-5	CCAAGCCCAATGTGCGCGACTCATTGC
MpOrf4v2-R36A-3	GCAATGAGTCGCGCACATTGGGCTTGG
MpOrf4v2-R236A-5	GGTGACCATGCGGTTCCGCTG
MpOrf4v2-R236A-3	CAGCGGAACCGCATGGTCACC

Table C2: Plasmids used

Protein	Plasmid name	Parent Vector	Tag	MW (kDa)	Reference
CitA	pEMpOrf4v2	pET-24a	C-His6	30.7	This work
CitA-S122A	pEMpOrf4v2-S122A	pET-24a	C-His6	30.7	This work
CitA-R36A	pEMpOrf4v2-R36A	pET-24a	C-His6	30.7	This work
CitA-R236A	pEMpOrf4v2-R236A	pET-24a	C-His6	30.7	This work
PksCT SAT-KS-MAT	pEMpPksCT-SKM3	pET-24a	C-His6	144.0	Appendix A
PksCT PT	pEMpPksCT-PT	pET-24a	C-His6	41.5	Appendix A
PksCT ACP	pEMpPksCT-ACP	pET-24a	C-His6	15.0	Appendix A
PksCT CMeT	pEMpPksCT-CMeT	pET-24a	C-His6	46.6	Appendix A
PksCT CMeT-R	pEMpPksCT-CMeT-R	pET-24a	C-His6	92.7	Appendix A

Figure C1: SDS-PAGE of PksCT ACP, CitA, and CitA mutants.

Separation of purified PksCT ACP and CitA by SDS-PAGE (16%) to show purity, with 5 μL per lane at 2 mg/mL total protein. M: BenchMark ladder (Invitrogen), with 50 and 20 kDa bands labeled, 1: CitA, 2: CitA-S122A, 3: CitA-R36A, 4: CitA-R236A, 5: PksCT ACP.

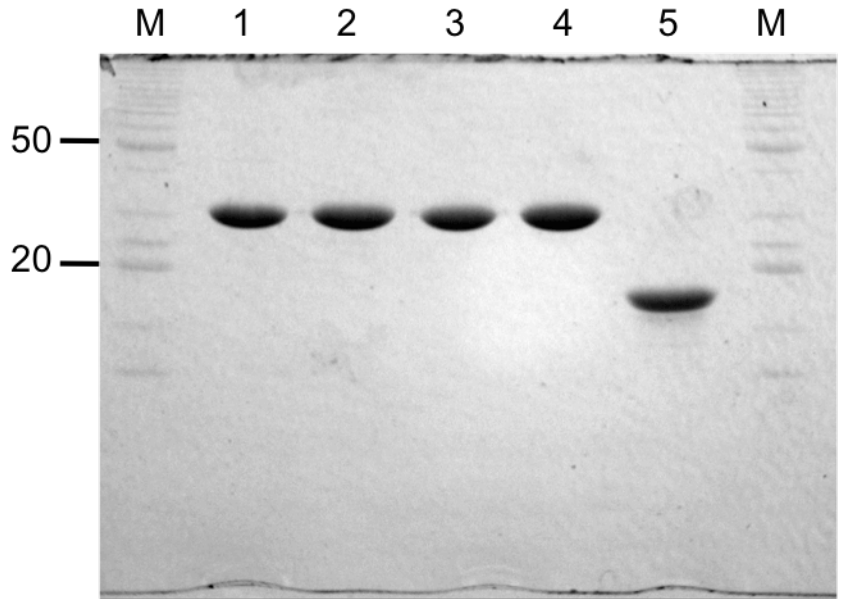
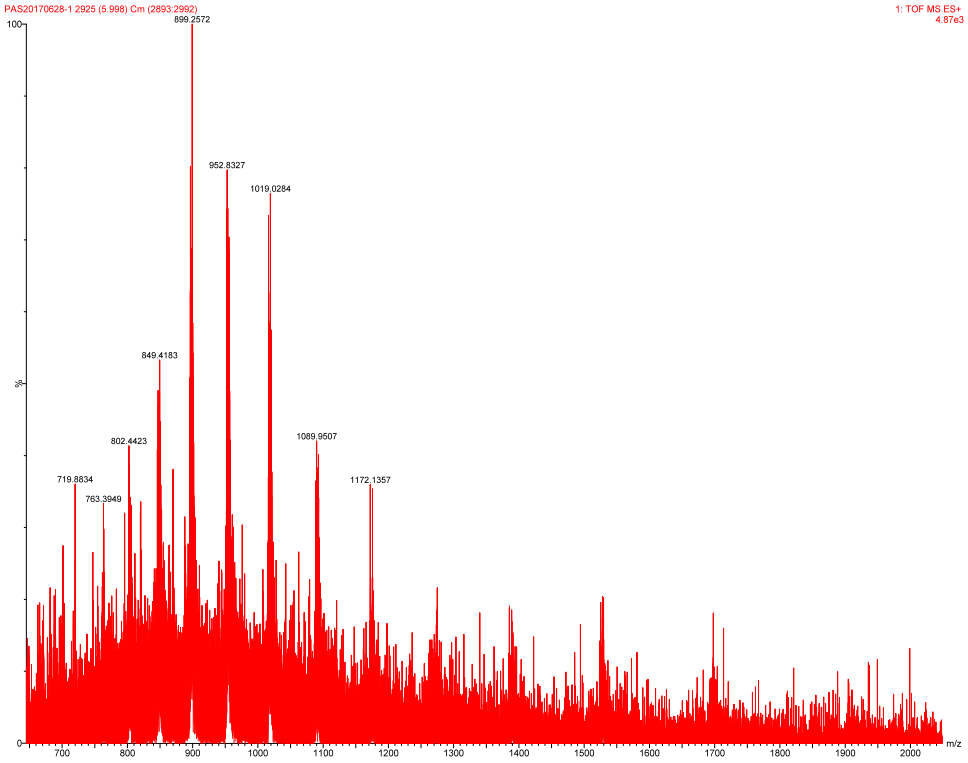
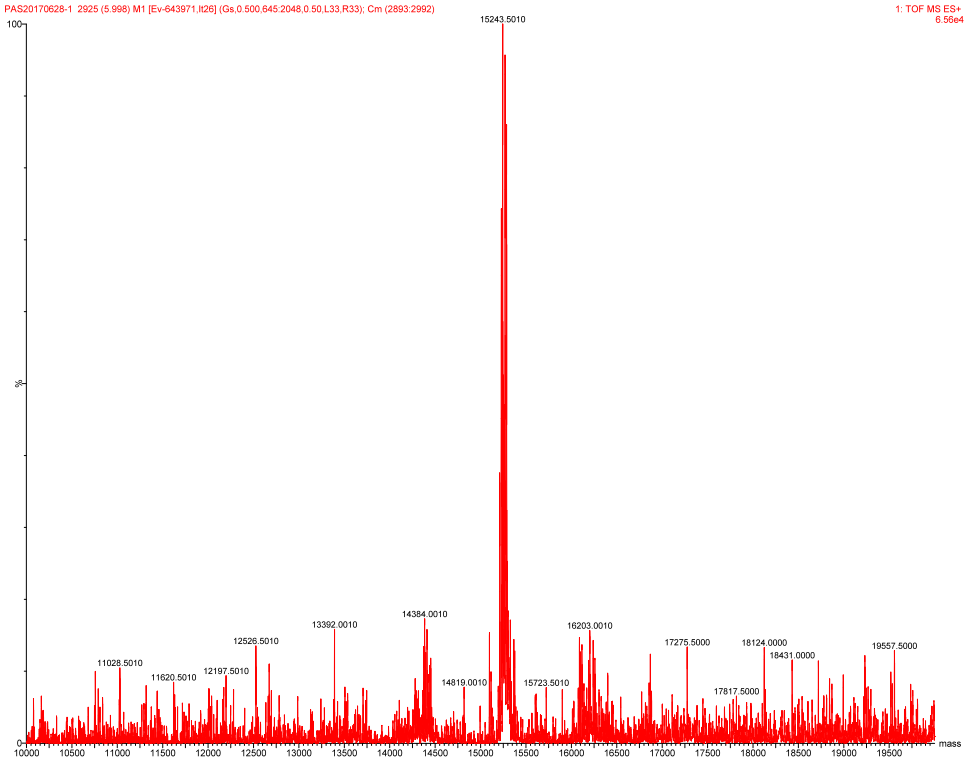


Figure C2: UPLC-ESI-MS data for apo-ACP_{CT}.

C2: Detected ions.



C2: Deconvoluted intact mass.



C2: Zoomed intact mass.

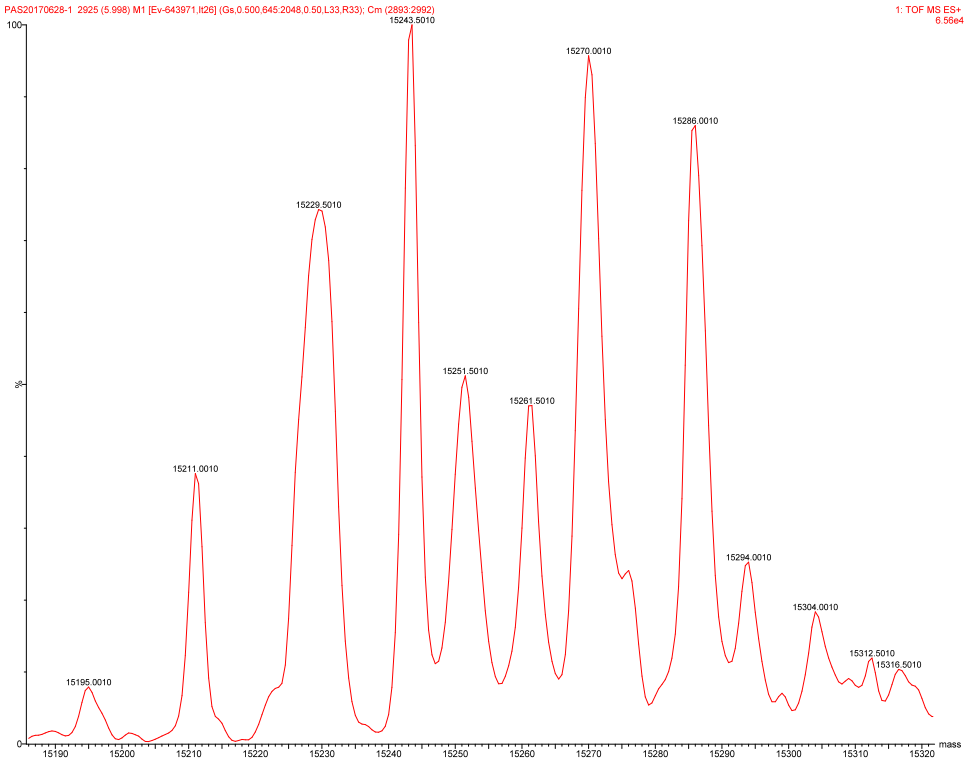
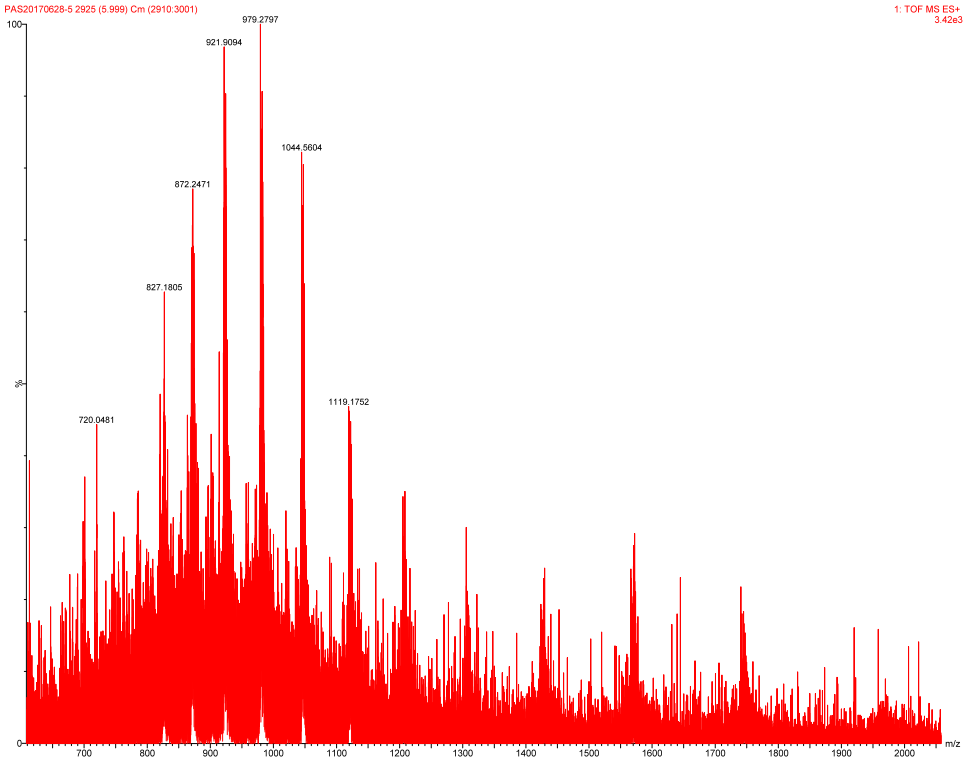


Figure C3. UPLC-ESI-MS data for Mal-holo-ACP$_{CT}$

C3: Detected ions.



C3: Deconvoluted intact mass.

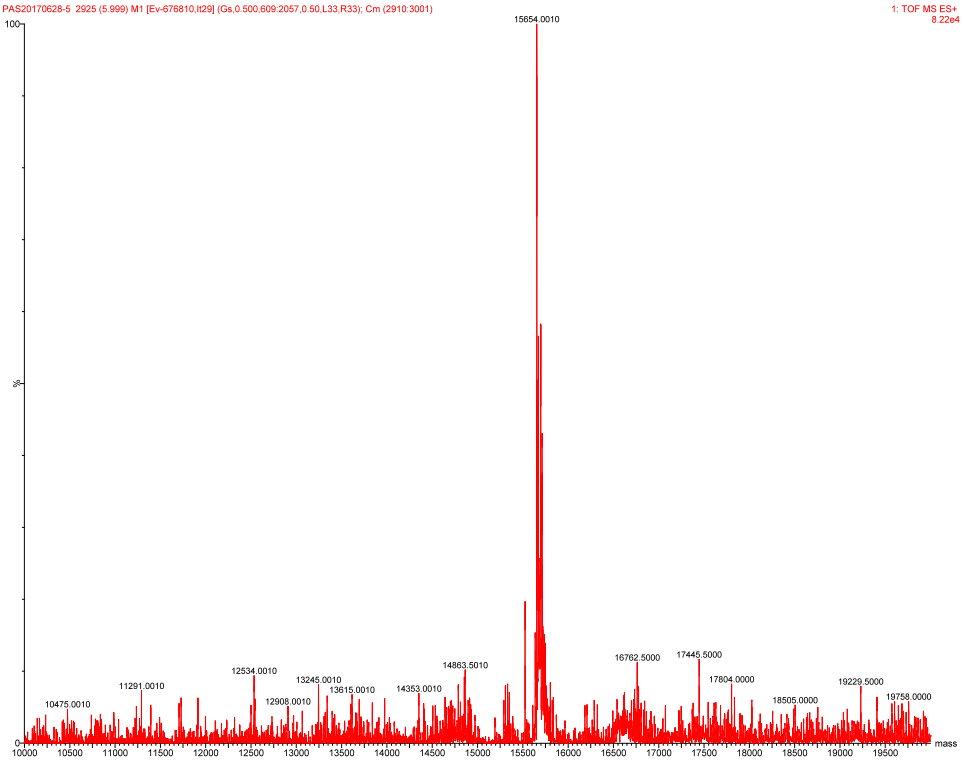
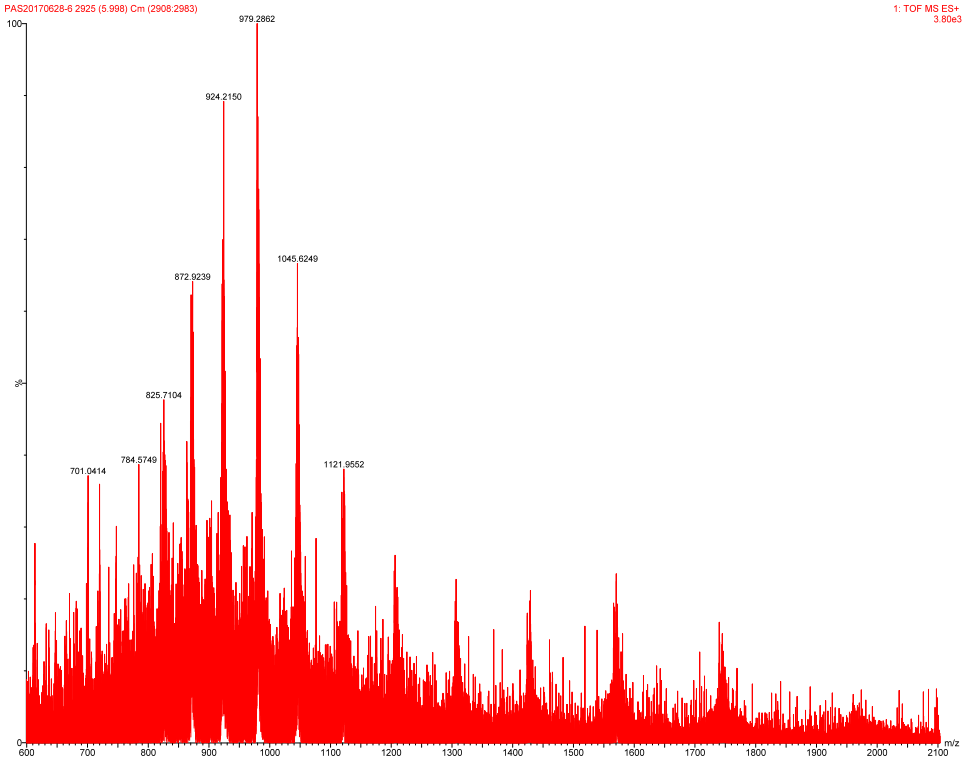


Figure C4. UPLC-ESI-MS data for Acac-holo-ACP$_{CT}$

C4: Detected ions.



C4: Deconvoluted intact mass.

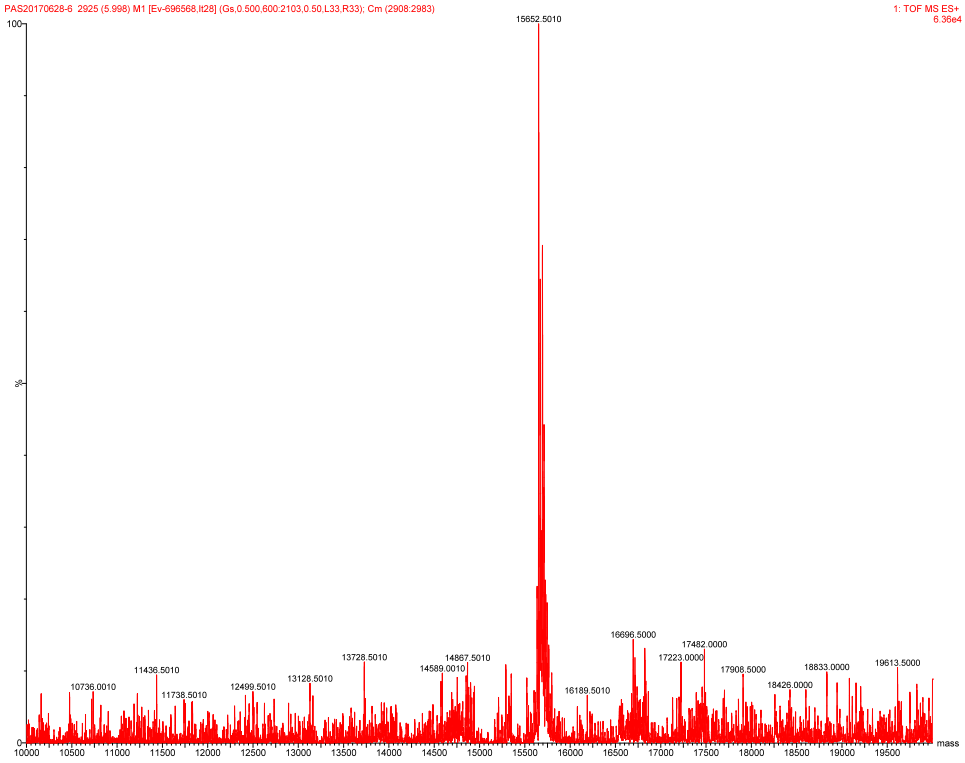
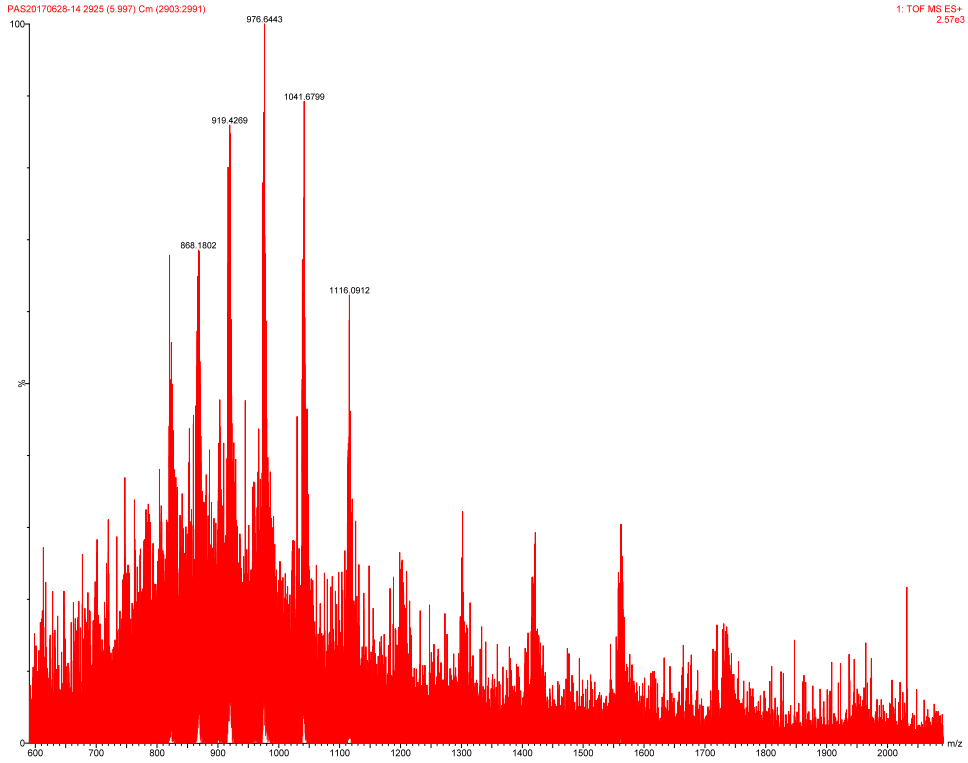


Figure C5. UPLC-ESI-MS data for Mal-holo-ACP$_{CT}$ + CitA

C5: Detected ions.



C5: Deconvoluted intact mass.

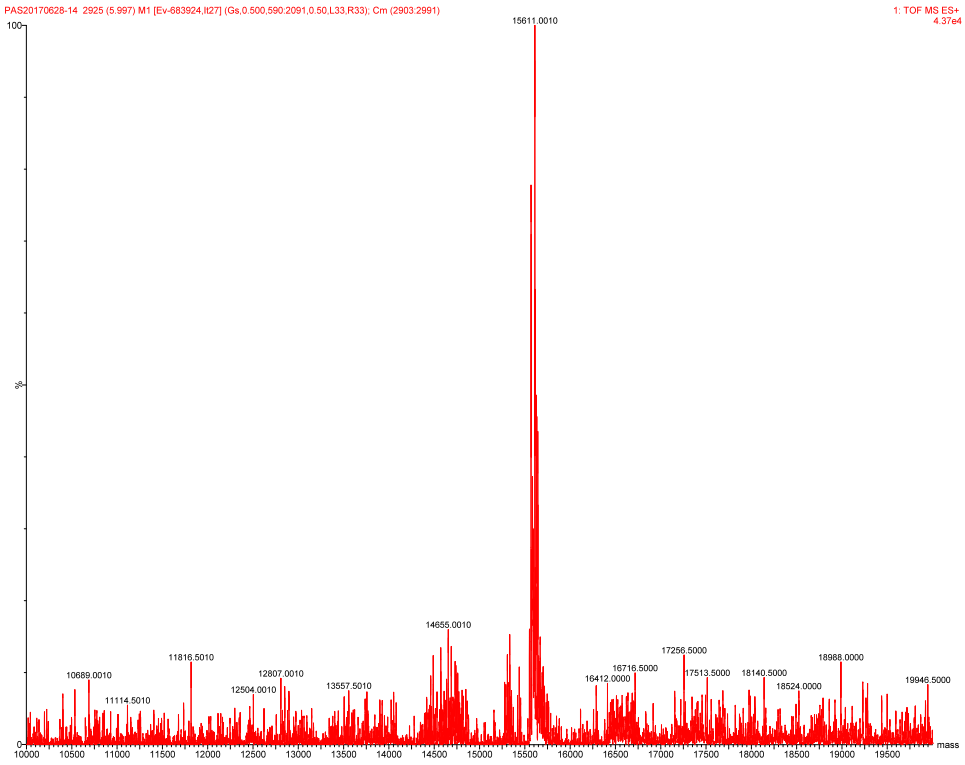
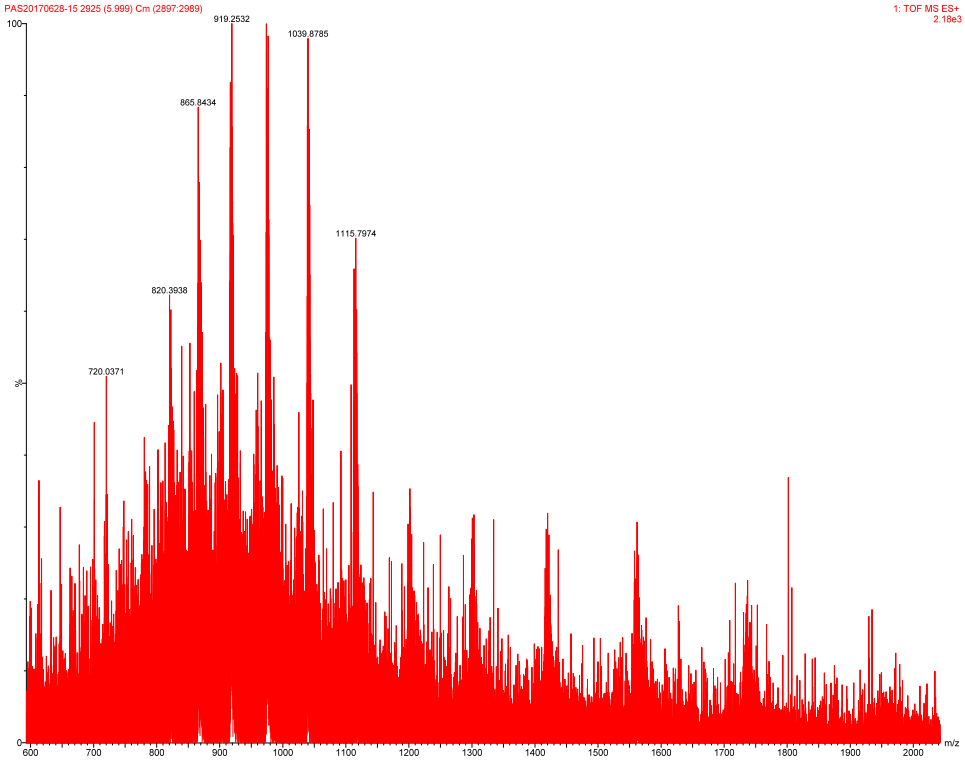


Figure C6. UPLC-ESI-MS data for Acac-holo-ACP_{CT} + CitA

C6: Detected ions.



C6: Deconvoluted intact mass.

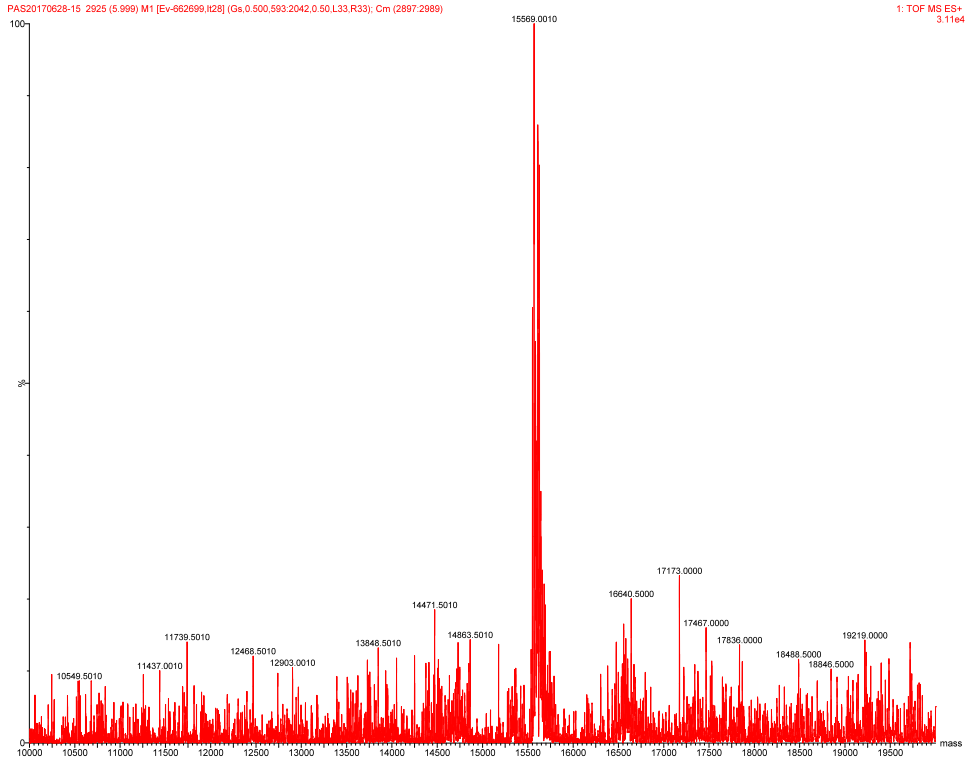


Figure C7: UPLC-ESI-MS traces for Mal-holo-ACP_{CT}/Acac-holo-ACP_{CT} + CitA

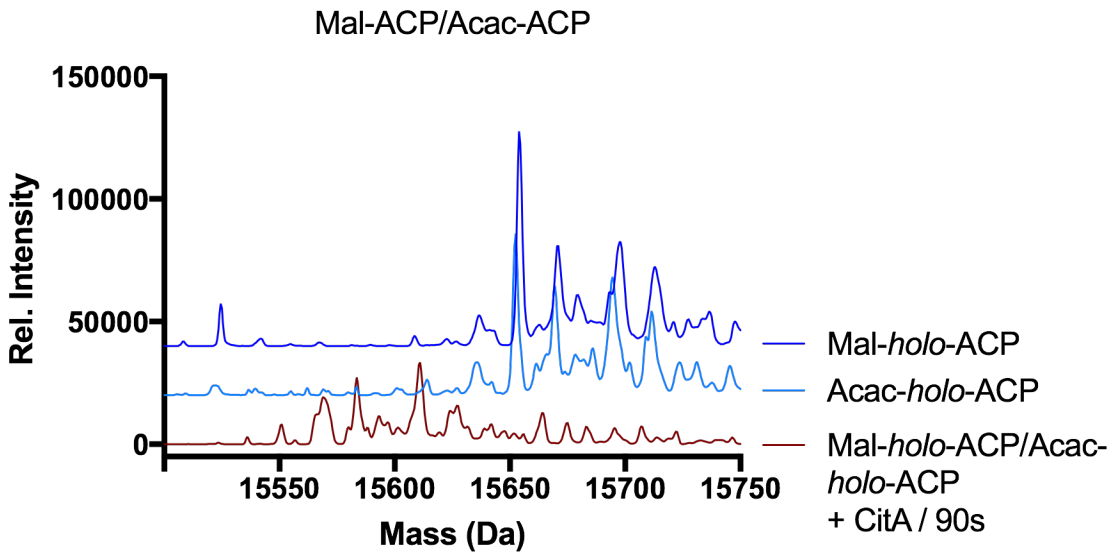
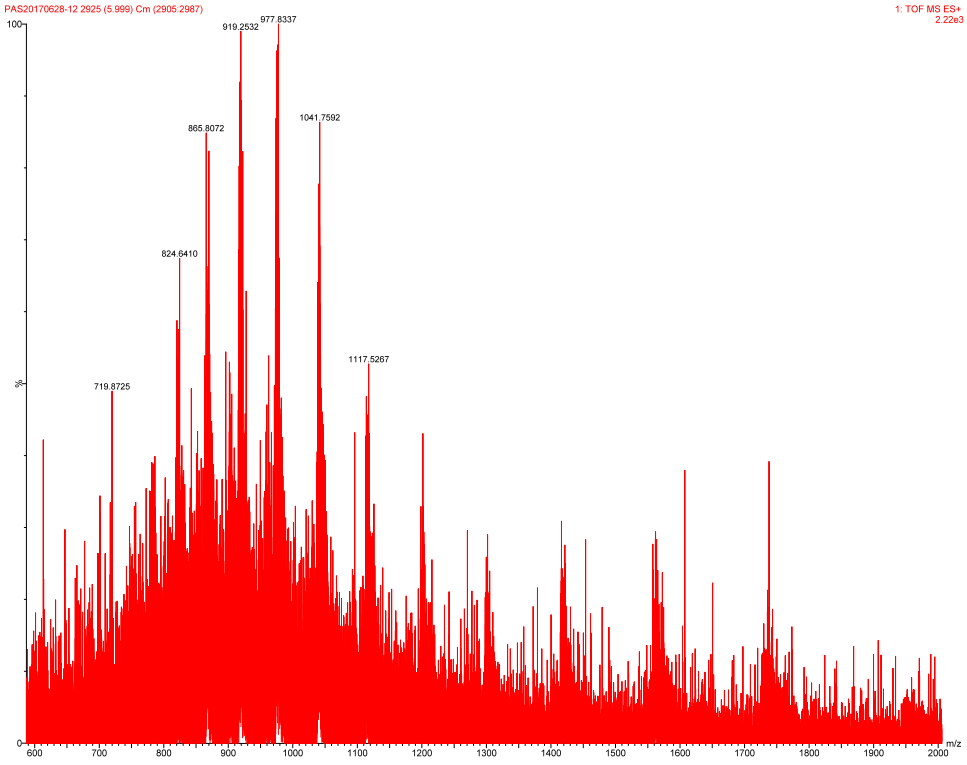
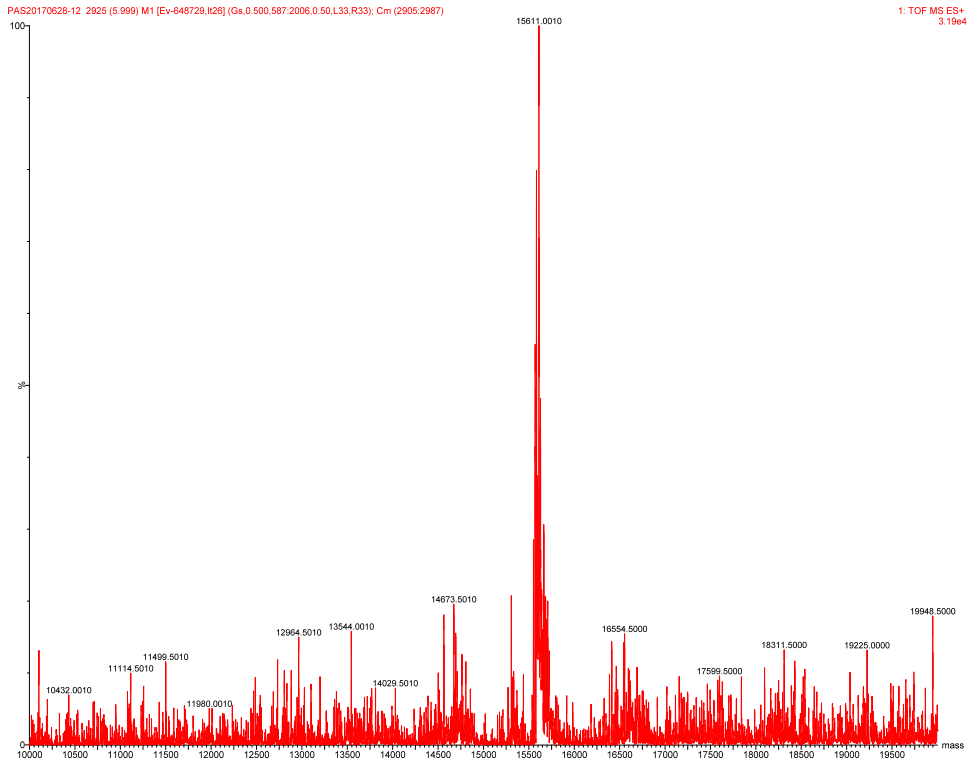


Figure C8: UPLC-ESI-MS data for Mal-holo-ACP$_{CT}$/Acac-holo-ACP$_{CT}$ + CitA

C8: Detected ions.



C8: Deconvoluted intact mass.



C8: Zoomed intact mass.

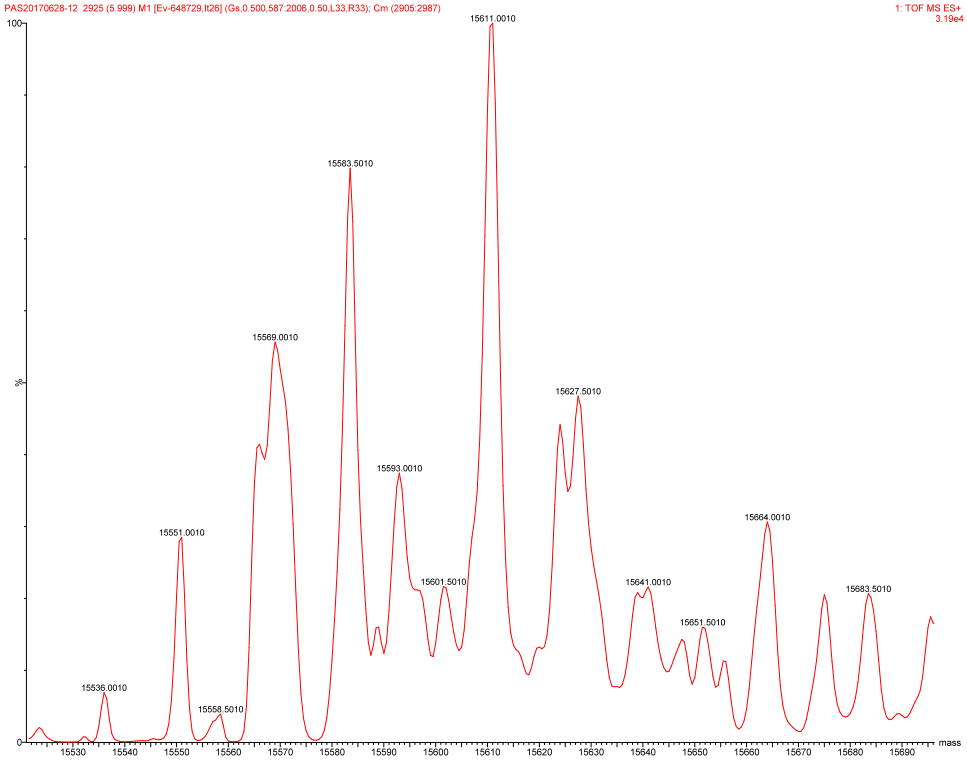
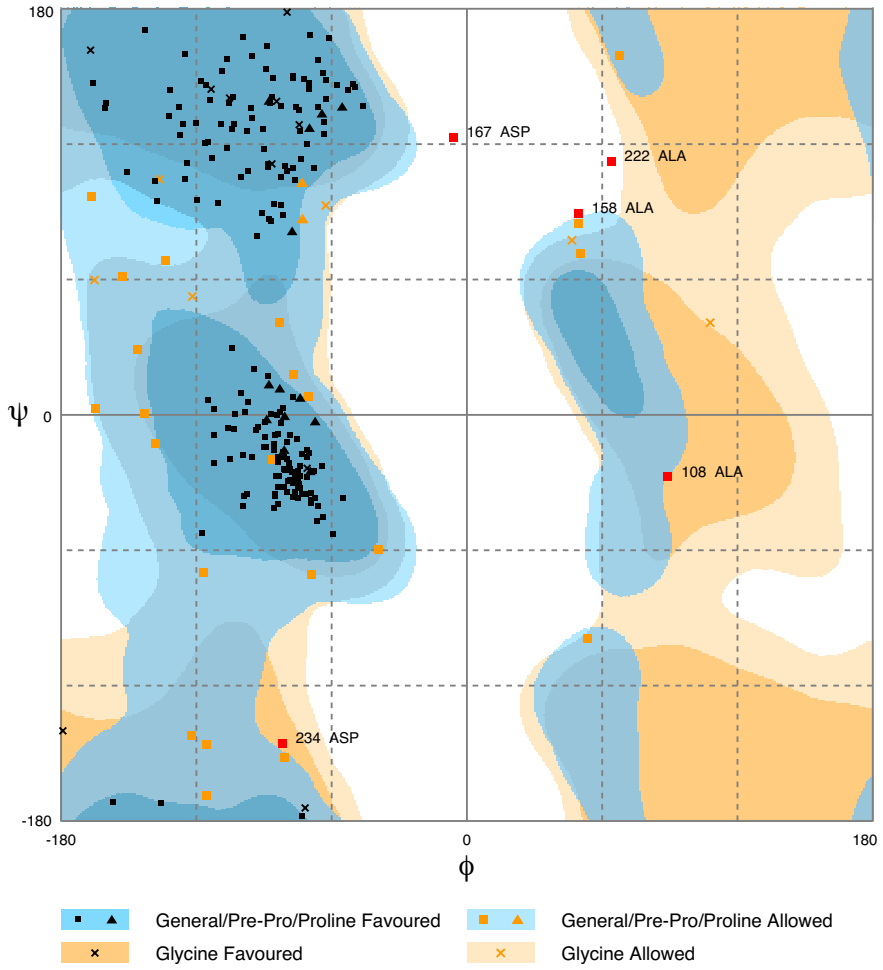


Figure C9: Rampage analysis of CPHmodel v3.2 homology model.



Number of residues in favoured region (~98.0% expected)	: 207 (85.5%)
Number of residues in allowed region (~2.0% expected)	: 30 (12.4%)
Number of residues in outlier region	: 5 (2.1%)

RAMPAGE by Paul de Bakker and Simon Lovell available at http://www-cryst.bioc.cam.ac.uk/rampage/

Please cite: S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson & D.C. Richardson (2002)
Structure validation by Cα geometry: ϕ/ψ and Cβ deviation. Proteins: Structure, Function & Genetics. **50**: 437-450

Cirriculum Vitae

Philip A. Storm was born on April 22nd, 1988, to Dean and Dorothy in Boston, Massachusetts. He and his family moved to Fairfax City, Virginia, and then to Rome, Italy, as he entered grade school. The two years spent there were some of the most influential and exhilarating – from standing on the U.S.S. George Washington aircraft carrier, to the pits of the Coliseum, to the Sistine Chapel. He returned to Virginia, where he attended St. Leo's Elementary School and then Bishop O'Connell High School, in Arlington, Virginia. During this time he was able to hike in the Rockies, sail in the Bahamas and Florida Keys, and gaze upon the Cascades. For his undergraduate education, he matriculated at Mr. Jefferson's University, the University of Virginia, in Charlottesville. After taking too many chemistry classes, he thought he should become a professional. He began his research with Prof. Mario Geysen and then with Prof. John Bushweller at U. Va., and a summer fellowship with Prof. Mark Braiman at Syracuse University. After graduating in May, 2010, he joined the Chemistry-Biology Interface program at Johns Hopkins University, in Baltimore, Maryland. At the end of his first year, he joined the laboratory of Prof. Craig Townsend. Under his guidance, Philip gained expertise in natural product biosynthesis, particularly in the area of fungal polyketides.